



Jouer avec le faux. Recherches sur les processus mentaux à l'oeuvre dans la lecture des textes de fiction.

Brian Hill

► To cite this version:

Brian Hill. Jouer avec le faux. Recherches sur les processus mentaux à l'oeuvre dans la lecture des textes de fiction.. Philosophie. Université Panthéon-Sorbonne - Paris I, 2006. Français. NNT : . tel-00108765

HAL Id: tel-00108765

<https://theses.hal.science/tel-00108765>

Submitted on 23 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris I Panthéon-Sorbonne
U.F.R. de Philosophie



THÈSE

présentée publiquement le 3 juillet 2006

par

Brian Hill

en vue de l'obtention du

Docteur de l'Université de Paris I

en

Philosophie

Jouer avec le faux

Recherches sur les processus mentaux à l'œuvre
dans la lecture des textes de fiction

Directeur de thèse :

M. Jacques Dubucs Directeur de Recherche au CNRS, Directeur de
l'IHPST

Jury :

M. Gregory Currie Professeur à l'Université de Nottingham, Directeur
de la Faculté des Arts

M. Philippe Mongin Directeur de Recherche au CNRS, Professeur au
HEC

M. Roger Pouivet Professeur à l'Université de Nancy 2

M. Jean-Marie Schaeffer Directeur de Recherche au CNRS, Directeur
d'Études à l'EHESS, Directeur du CRAL

M. Robert Stalnaker Professeur au Massachusetts Institute of Technology

Remerciements

Je tiens à remercier avant tout Jacques Dubucs. Le soutien, l'attention et la confiance qu'il m'a témoignés durant ces années de thèse, et l'année qui les a précédée, sont inestimables. Il lui a fallu une patience considérable, une vision admirable, pour ne pas mentionner la courage qui est la sienne, pour accepter en DEA un étudiant qui n'est ni francophone ni philosophe d'origine. Je me sens particulièrement fortuné d'avoir pu bénéficier de cet esprit large, qui est d'une rareté regrettable.

Je souhaite également remercier Gregory Currie, Philippe Mongin et Robert Stalnaker d'avoir accepté de faire partie du jury de cette thèse, et particulièrement Roger Pouivet et Jean-Marie Schaeffer qui ont accepté la charge supplémentaire de lire la version préliminaire de ce travail.

Je remercie ceux qui ont consacré quelques moments à des parties de ce travail, dont Arnaud Plagnol et Richard Bradley. J'ai une dette particulière à l'égard des membres du jury de la pré-soutenance de cette thèse : Roger Pouivet, dont les conseils et les commentaires se sont avérés grandement utiles, Gabriel Sandu, dont les remarques critiques ont motivé une partie du travail, et notamment Jean-Marie Schaeffer, dont les commentaires, l'enthousiasme et l'encouragement m'ont beaucoup inspiré. J'ai eu une chance particulière de les connaître, et d'avoir eu des discussions extrêmement intéressantes et fécondes avec eux.

J'ai bénéficié, pendant ces dernières années, de l'environnement, intellectuel aussi bien que social, de l'IHPST. Je tiens à remercier des membres de l'institut, et notamment les doctorants, non seulement pour leur suggestions et leurs conseils, qu'elles soient philosophiques, techniques, ou linguistiques, mais également pour leur amitié. Tout particulièrement Denis Bonnay, Cédric Paternotte, Henri Galinon, Amélie Gheerbrant, Julien Boyer, Mikael Cozic, Paul Egré, Peggy Cardon, Mark van Atten et Wioletta Miskiewicz ont apportés, chacun à sa façon, de l'aide ou de l'inspiration à la réalisation de ce travail.

Last, but not least, je ne sais pas comment exprimer mes remerciements

auprès de ma famille et de mes amis, dont l'affection indéfectible et les nombreux autres soutiens petits mais cruciaux ont eu une importance que personne ne saurait quantifier. Je tiens à consacrer un mot particulier pour la patience équivoque, l'humour particulière, et l'érudition fabuleuse de Francesca, qui ont rendu plus supportables ces derniers mois.

Paris, mai 2006

Avertissement

Malgré la diversité des sujets traités – la fiction, la psychologie de sa lecture, les contrefactuels – un des points importants de la thèse présentée dans les pages qui suivent est son unité. Comme il sera souligné à plusieurs reprises, il s’agit d’aborder ces domaines différents dans une même perspective. Afin de saisir l’unité de cette approche, il conviendrait donc de s’intéresser spécifiquement aux différents domaines discutés. Néanmoins, j’espère que les considérations spécifiques à chaque domaine pourront être intéressantes en elles-mêmes. Pour cette raison, je me propose ici d’ébaucher le plan général de la thèse, afin d’aider ceux qui ne souhaitent pas la lire en totalité, mais seulement se concentrer sur certaines parties.

Le plan de la thèse est comme suit :

Introduction générale à la problématique de la fiction et des contrefactuels ;
brève esquisse de l’approche proposée.

Partie première portant sur les processus mentaux impliqués dans la lecture des textes de fiction.

Ch 1 (introduction à la problématique générale et considération d’autres théories) :

- établissement de la problématique (§1.1)
- considération générale des approches traditionnelles de cette problématique et introduction à l’approche proposée ici (§1.2)
- préliminaires de la théorie de l’esprit (§1.3)
- discussion critique des théories psychologiques actuelles de la fiction (§§1.4 et 1.5).

Ch 2 (présentation d’une théorie psychologique de base qui permet une analyse approfondie de la lecture de la fiction) :

- outils de base 1 : introduction à la modélisation formelle de l’état instantané de la conscience (§§2.1 et 2.2)

- outils de base 2 : discussion et prise en compte de la dynamique de la conscience (§2.3)
- comment retrouver les États Mentaux, tels les croyances et les émotions, dans le cadre proposé (§2.5).

Ch 3 (application de la théorie psychologique proposée à la fiction)

- théorie des compétences du lecteur par rapport au caractère fictionnel du texte (§3.1)
- remarques relatives aux « concrétisations » des propos du texte par le lecteur, et aux « infiltrations » des prédicats, personnages ou thèmes des textes de fiction dans la réalité (§3.2)
- diagnostic des faiblesses des États Mentaux proposés comme spécifiques à la fiction, tel le *faire comme si* (§3.3).

Partie seconde proposant une analyse des contrefactuels

Introduction (discussion générale de la problématique des contrefactuels, du rapport à la fiction et à sa problématique, et de l’usage de ce rapport dans cette thèse ; discussion des théories de Lewis relatives à la fiction et aux contrefactuels ; présentation de l’approche développée dans cette partie).

Ch 4 (établissement de la problématique des contrefactuels) :

- introduction aux contrefactuels, à leur rapport à la fiction, et au problème qu’ils posent (§4.1)
- établissement d’une liste de défis pour une théorie de contrefactuels : rendre compte d’un certain nombre de propriétés « intuitives » des contrefactuels (repérées dans §4.2) ; s’occuper des aspects qu’on laisse souvent de côté sous le nom de la « pragmatique » (§4.3).

Ch 5 (ébauche d’une théorie de contrefactuels) :

- exposition de la théorie, procédant par étapes (§5.1)
- évaluation de la théorie relativement aux défis repérés dans le Chapitre 4 (§5.2)
- discussion du rapport de la théorie proposée, d’une part à la distinction normalement admise entre la sémantique et la pragmatique, et d’autre part à la question de la « logique » des contrefactuels (§5.3).

Conclusion résumé des problèmes communs à la fiction et aux contrefactuels, et de l’opposition d’approches possibles de ces problèmes ; rapprochement entre cette problématique et celle de la potentialité au sens général, et

esquisse de quelques pistes de recherche pour étendre le travail dans cette direction.

Annexe discussion de trois interprétations des règles logiques dans le cadre théorique proposée dans l'analyse des contrefactuels, et de la plausibilité des règles traditionnelles de la logique non monotone sous les différentes interprétations.

Toujours dans le but de clarifier la démarche de la thèse, j'ai placé au début de chaque paragraphe une petite vignette qui en résume les points principaux. Dans des paragraphes faisant référence à des définitions ou à des résultats techniques (notamment dans le Chapitre 5), j'ai privilégié le langage courant dans l'essentiel du texte, en rassemblant les détails techniques à la fin du paragraphe.

On s'est pensé naïf, on s'est avéré compliqué

J. Joyce

Introduction

Que c'est drôle, la carte d'Afrique ! Pays séparés par des lignes droites et simples, comme si la simplicité géométrique était le critère important d'une frontière entre nations. Drôle, mais, comme on le sait trop bien, finalement triste. Une ligne rude, simplissime, posée de manière abstraite depuis une vue lointaine ne convient nullement ; il faut plutôt aller sur le terrain, au niveau local voire concret, pour repérer ses particularités et ses complexités, géographiques, politiques, sociales, culturelles. Seulement ensuite on peut faire la synthèse de ces observations locales et relativement fidèles aux faits, pour en faire le tri mûr, pesé, correct si l'on peut dire. Certes, cette dernière démarche est beaucoup plus difficile, lente et exigeante ; n'empêche qu'il faut condamner ceux qui se contentent de la première. À plus forte raison s'il s'agit de philosophes.

En effet, en ce qui concerne les objets de cet essai – la fiction, les contre-factuels (des conditionnels à l'antécédent actuellement faux) et notamment leur rapport à la réalité – la démarche simplissime et rude a trop longtemps dominé. On pose la fiction et la réalité comme deux notions, nations, voire mondes, s'opposant l'un l'autre, alors que le rapport entre elles est en vérité beaucoup plus complexe et subtil. Ainsi les complexités de leur rapport échappent. De la même manière, on raccroche le contrefactuel à un monde possible ou situation irréaliste qui se tient face à la réalité, alors que contrefactuel et réalité sont notoirement imbriqués. L'opposition simple tend à écraser ces complexités vives, à tel point qu'elles se trouvent dédaignées voire tout simplement ignorées.

Il faut contester cette démarche, en revendiquant l'approche plus précise et fidèle qui consiste à partir de la complexité et des détails du rapport entre fiction et réalité, ou entre contrefactuel et réalité, pour en venir seulement après aux grandes lignes de ce rapport. Il s'agit d'entrer directement sur le terrain où la fiction et la réalité (respectivement le contrefactuel et la réalité) sont enchevêtrées, dans les localités où elles se confrontent et s'infiltrant, pour ensuite rassembler les données des différents lieux particuliers. Si la fiction et la réalité, ou le contrefactuel et la réalité, n'impliquent pas un clivage net d'un espace,

mais plutôt un enjeu et un chevauchement d'aspects, d'inclinations et d'aires diverses, ce sont les détails du paysage qu'il importe de comprendre, et cela ne pourrait se faire qu'à partir de contrées particulières. Le but de cet essai est de proposer une telle approche de la fiction et des contrefactuels.

A première vue, on pourrait être étonné par la mise en rapport de notions aussi disparates que la fiction et les contrefactuels. La communauté entre elles est cependant profonde, tenant pour l'essentiel à leurs rapports respectifs à la réalité. En premier lieu, aussi bien la fiction que le contrefactuel ont affaire à ce qui n'est pas strictement réel ou vrai – les événements fictionnels, les situations irréelles. Néanmoins, et en second lieu, il y a un certain sens de vérité ou de correction qui leur sont pertinents – il fait sens de parler de « vérités » selon un texte de fiction, ou de parler de la vérité ou la correction d'un contrefactuel. En un mot, la fiction et le contrefactuel *échappent à la réalité*, mais *cela n'implique pas que n'importe quoi se produise*. En troisième lieu, dans les deux cas, ce qui se produit *dépend dans une large mesure de la réalité* : ce qui est vrai selon un texte de fiction, ou le fait pour un contrefactuel d'être vrai ou non, dépend des faits *réels*, de la réalité. Le Londres des nouvelles de *Sherlock Holmes* a la même topologie que le Londres réel, même si ce n'est pas explicitement affirmé dans les nouvelles ; un contrefactuel « si tel événement historique n'avait pas eu lieu, alors . . . » est valide ou vrai en raison des faits réels qui le « soutiennent ». En dernier lieu, ces rapports avec la réalité *ne sont pas nécessairement unidirectionnels* : on suggère souvent que la fiction et les contrefactuels contribuent à la réalité, du moins à la réalité en tant que l'on la conçoit. On hérite d'Aristote l'idée que la fiction devrait viser à dire « le général »¹ : mais le « général » n'est-ce donc pas quelque chose de commun à la fiction et à la réalité ? Et donc, raisonne-t-on souvent, si dans une fiction est repérée, accentuée, voire même dévoilée une « généralité » peu connue ou reconnue dans la réalité, n'a-t-elle pas contribué en quelque manière à la conception de la réalité, sinon à la réalité elle-même ? De manière semblable, on sait bien que les contrefactuels jouent des rôles importants dans la considération de questions aussi diverses que la causalité, les lois scientifiques, le changement de croyances, c'est-à-dire des questions importantes ayant trait à la réalité.

On a donc bien raison de rapprocher la fiction et le contrefactuel, sous l'enseigne de la question qui leur est commune : *la question de leurs rapports respectifs à la réalité*. Cette question domine et motive cet essai.

Pour ceux qui sont habitués à l'exactitude, cette question pourrait paraître vague et peu précise. Or il faut rappeler que la réponse qui se trouve à la fin

¹Aristote, *Poétique*, 1451 b 1-10.

d'une recherche est déjà esquissée dans la question qui amorce cette recherche. Dans l'acte de préciser une question, on restreint son champ, son domaine, les cas d'intérêt, de telle sorte que la réponse se trouve plus sûrement et plus rapidement ; or cette réponse est marquée par les particularités des précisions portées, elle implique et repose sur les concepts et les astuces qui n'ont de sens que dans les domaines limités. En revanche, une question peu particularisée déborde ces spécificités ; elle souffre cependant de la pauvreté de concepts, d'astuces et d'outils qui sont disponibles à celui qui lui cherche une réponse. C'est ce deuxième genre de question dont s'occupe ce travail, et ce deuxième genre de réponse qu'il tente d'offrir. Il s'agit de la question du rapport entre fiction et réalité, et du rapport entre contrefactuel et réalité, dans sa généralité ; les éventuelles applications des conclusions aux domaines particuliers ayant un rapport avec la fiction ou les contrefactuels ne figureront pas parmi les considérations. Tout le travail baigne dans cet esprit général, et ce recul devant la spécificité : non seulement les concepts de fiction et de contrefactuels, mais également les discussions du mental et de la logique, auront ce caractère.

En conséquence de ce souci de généralité, on ne posera pas la question de l'*identité* de la fiction, ou de la *définition* des contrefactuels. Car, dans la mesure où les réponses à ce genre de question réduisent la fiction ou les contrefactuels à des facteurs provenant des domaines particuliers – en les caractérisant par des critères sémantiques, linguistiques, métaphysiques, ontologiques, pragmatiques ou autres –, elle ne pourrait que limiter l'étendue de l'enquête. Il sera donc nécessaire de travailler avec des notions immédiates, floues mais naturelles de fiction ou de contrefactuel, pour être sûr de les saisir dans toutes leurs complexités. Il ne sera donc pas question de se limiter aux textes de fiction particuliers, ou des œuvres d'une qualité particulière ; il s'agira de la fiction en tant que l'on la trouve dans le rayon « fiction » des librairies, et non de la littérature au sens de belles lettres. Si l'attention ici porte sur la fiction écrite et lue, ce n'est que pour simplifier le travail : les résultats devraient être tellement généraux qu'ils pourraient s'appliquer *mutatis mutandis* à d'autres dispositifs fictionnels. De manière pareille, la considération ne se bornera nullement aux contrefactuels en tant que phénomènes linguistiques ou en tant que simples constructions logiques, car elle leur vise dans une généralité plus grande. La fiction, les contrefactuels comme on les trouve *en pratique*, avec toutes leurs ambiguïtés, leurs influences, leurs effets, leurs emplois : dans ce travail, l'*effectivité* est l'ordre du jour.

Du même coup, il ne pourra être question de *supposer, spécifier ou soutenir aucun rôle ou fonction* à la fiction ou aux contrefactuels, car cela les inscrirait dans des champs particuliers. Aucune supposition, donc, sur l'importance, so-

ciale, morale ou autre de la fiction, ni sur l'importance des contrefactuels par rapport à la question des lois, de la causalité, de la croyance, par exemple. Par conséquent, aucun droit de laisser entrer dans les analyses de la fiction et des contrefactuels des aspects tenant aux rôles particuliers qu'ils peuvent remplir, aucune possibilité, par exemple, d'analyser le contrefactuel en termes de son rapport à la causalité.

Enfin, quant aux propriétés « profondes » de la fiction et des contrefactuels, qu'elles soient logiques, métaphysiques, ontologiques, épistémologiques ou autres, en s'en occupant de l'une ou l'autre d'entre elles, et en se permettant des concepts spécifiques à ces domaines, on renonce à la généralité du questionnement. La recherche menée ici se privera donc de tout concept spécifiquement métaphysique, ontologique, épistémique, elle évitera tout questionnement qui appartient à aucun de ces domaines, elle sera complètement *neutre* par rapport à ces débats. L'existence et les propriétés des « êtres fictionnelles », le teneur métaphysique de contrefactuels, la nature – réaliste, instrumentaliste, anti-réaliste – des Etats Mentaux et de notre connaissance à leur égard : les considérations et les conclusions dans les pages qui suivent seront consistantes avec toute position relative à ces questions, quoiqu'elles puissent être intéressantes à celui qui s'en occupent. Car, s'il fallait qualifier le genre de question qui est posée ici, elle serait *méthodologique* : *quelle est la meilleure manière de penser ou de se représenter la fiction ou les contrefactuels pour comprendre leurs rapports effectifs avec la réalité ?* La conception que l'on se fait de la fiction ou des contrefactuels en tant qu'ils se rapportent à l'être, à la connaissance, au langage, *pourrait* tirer profit de l'enquête sur la meilleure manière de les concevoir en général ; les détails appartiennent aux domaines spécifiques et à leurs spécialistes.

On arrive à mieux comprendre la démarche suivie ci-dessous, et la question qui la motive, par le biais de la comparaison avec celle qu'entreprend Lewis dans un article fameux². Lewis analyse la question de la *vérité fictionnelle* – les vérités autorisées par un texte de fiction – en termes de contrefactuels : à quelques détails près, « *X* est vrai selon la fiction *f* » si et seulement si le contrefactuel « si *f* était raconté comme vrai, alors *X* serait vrai » est vrai. Ce contrefactuel est à son tour analysé en termes de « mondes possibles » : il est vrai si *X* est vrai dans tous les mondes possibles les plus proches où *f* est raconté comme vrai. Or, quelles que soient les intuitions et les avantages *métaphysiques* de cette mise en jeu des mondes possibles, ils n'ont aucune valeur quant à la question posée à l'égard des particularités *effectives* et *pratiques* de la fiction et

²Lewis (1978).

des contrefactuels ; ce qui importe, en revanche, c'est le sens de la « similitude » ou « proximité » entre mondes qui était cette analyse. La notion de similitude employée dans l'analyse est cependant très mal comprise, dépendant, selon le mot de Lewis³, de « la pratique et du contexte ». Par conséquent, son analyse n'est satisfaisante *ni* comme réponse à la question du rapport entre contrefactuels et réalité *ni* comme réponse à la question du rapport entre fiction et réalité. *Méthodologiquement*, cette mise en jeu de mondes possibles ne s'avère pas féconde.

Cette thèse prend comme point de départ la communauté entre la fiction et les contrefactuels qui sous-tend la proposition de Lewis. Puisqu'il n'y a aucune question de « fonder » la fiction et les contrefactuels l'un sur l'autre ou de « réduire » l'un à l'autre, on ne reconnaît aucune *priorité logique* entre fiction et contrefactuels. Néanmoins, l'ordre de considération des deux domaines implique un ordre *heuristique* entre fiction et contrefactuels, dans la mesure où l'une sera abordée avant l'autre. Cet ordre est l'inverse de celui de Lewis : là où Lewis part d'une analyse des contrefactuels et mobilise le rapport avec la fiction pour « induire » une analyse de la vérité fictionnelle, il s'agira ici de développer d'abord une théorie relative à la fiction et la vérité fictionnelle pour ensuite « s'inspirer » de cette théorie pour proposer une analyse des contrefactuels. Plus précisément, il s'agira de développer *initialement* une théorie de la *lecture de la fiction*, c'est-à-dire une théorie de la *psychologie du lecteur* du texte fictionnel ; ce sera le but de la première partie de ce travail. *Ensuite*, en s'appuyant sur le rapport étroit, montré ailleurs⁴, entre la vérité fictionnelle et l'acte de lecture, on saura tirer des conséquences pour la question de la vérité fictionnelle. *Enfin*, dans la seconde partie, une théorie des *contrefactuels* qui s'inspire de la théorie de la lecture de la fiction et de la vérité fictionnelle sera proposée. En vérité, la démarche relative à la fiction et celle relative aux contrefactuels, ainsi que les théories auxquelles elles aboutissent, sont tout à fait parallèles.

La recherche sur la lecture de la fiction aussi bien que celle relative aux contrefactuels commence par un moment préliminaire et critique, servant à développer les intuitions relatives à leurs propriétés effectives et à questionner les théories déjà existantes de la lecture de la fiction et des contrefactuels (Chapitres 1 et 4). Ces théories se bornent le plus souvent à proposer une dualité simple : dans le cas de la lecture de la fiction, une dualité d'attitudes mentales – la croyance d'un côté, une attitude d'« imagination » ou de *faire comme si* relative aux énoncés fictionnels de l'autre ; dans le cas des contrefactuels, comme

³Lewis (1981, p87 (§5)). Voir aussi 1973b, §4.2

⁴Hill, *Quelques caractéristiques de la fiction*, Mémoire de DEA (2003), Université Paris I, §VIII.

on l'a vu chez Lewis, une dualité de mondes – monde réel d'un côté, monde « contrefactuel » de l'autre. Elles n'arrivent pas cependant à rendre compte de la complexité des attitudes envers la fiction et de leurs rapports aux attitudes ordinaires, et pas non plus à rendre compte de la particularité des contrefactuels et de leur rapport aux vérités réelles. En philosophie comme en politique, les distinctions simples et à première vue attirantes sont souvent trop naïves, n'offrant pas de vraie compréhension des subtilités, en l'occurrence du rapport entre les croyances du lecteur et ses attitudes de *faire comme si* à des moments de la lecture, et du rapport entre les contrefactuels et les vérités réelles.

Pour emprunter la voie plus dure mais finalement plus féconde mentionnée ci-dessus, il s'agit *en premier lieu* d'aborder les questions dans des *sites* ou *localités particulières*. Dans le cas de la lecture de la fiction, les « sites » appropriées pour penser les attitudes du lecteur à l'égard de la fiction et de la réalité, aussi bien que le rapport entre elles, ont affaire à des moments différents de la lecture ; en fin de compte, il s'agira des *états instantanés de la conscience* à ces moments où les attitudes à l'égard des propos du texte fictionnel et de la réalité extérieure se rencontrent et se trouvent en rapport (Chapitre 2 et 3). Dans le cas des contrefactuels, les « sites » appropriées fournissent les moyens pour évaluer les contrefactuels, à la façon de la relation de similitude posée par Lewis ; il s'agira en vérité des *contextes* en vigueur à ces moments (Chapitre 5). *Dans un second temps*, après avoir fait les « observations du terrain », il s'agit de faire la synthèse de ces observations pour aboutir à une vue plus ou moins globale. Dans le cas de la psychologie de la lecture, où les sites se distinguent temporellement, la question de la mise en rapport des différentes sites se ramène essentiellement à celle de leur *dynamique*, plus précisément, de la dynamique de la conscience. De manière semblable, dans le cas de contrefactuels, où les différents contextes s'enchaînent eux aussi dans le temps, la question devient celle de la *dynamique du contexte*, ou, pour l'exprimer en termes lewisiens, de la *dynamique de la relation de similitude*.

Cette démarche propose donc une *reconfiguration* du rapport entre fiction et réalité (en l'occurrence, en tant que ces termes apparaissent dans la psychologie du lecteur), du rapport entre contrefactuel et réalité, et des questions de ces rapports. Pour penser ces rapports de manière féconde, pour poser proprement les questions de ces rapports, il faut les factoriser en deux. Premièrement, il y a la question *instantanée* : du rapport entre les attitudes du lecteur à des instants particuliers ; du rapport entre les contrefactuels et les vérités réelles à des instants particuliers. Deuxièmement, il y a la question *diachronique* des changements du rapport d'un instant à l'autre, ou plus généralement des propriétés

de la *dynamique du rapport*. Dans le cas de la fiction, c'est la question de la dynamique du rapport entre fiction et réalité qui constitue le cœur épineux mais crucial de la question du rapport *effectif* ou *pratique* entre fiction et réalité. Quelle compréhension pourrait-on escompter du fait pour le lecteur de croire (explicitement), à la fin de sa lecture, que tel ou tel est vrai selon le texte, si ce n'est pas la compréhension de la manière dont la suite des attitudes à des moments antérieurs et des rapports entre ces attitudes, l'a « mené » à cet état ? Également dans le cas de contrefactuels, c'est la question de la dynamique du contexte, ou si l'on préfère de la relation de similitude, qui est au centre du problème posé par les contrefactuels. Si l'on évalue ce contrefactuel comme vrai, et si cette évaluation dépend des caractéristiques du contexte, la question des contrefactuels est : comment est-on venu à un contexte possédant de telles caractéristiques ? Une façon plus convenable de poser une question ancienne, une manière plus apte de figurer le champ de recherche : cette *conception* de la fiction, des contrefactuels, et du problème qu'ils posent est l'apport majeur de cet essai.

En outre, cette conception offre une compréhension simple et quasi-immédiate d'un certain nombre de propriétés ou phénomènes relatifs à la fiction et aux contrefactuels qui étaient antérieurement problématiques. D'un côté, la caractérisation des états instantanés de la conscience et de leur structure débouche par exemple sur une théorie de l'immersion fictionnelle comme absence du fait qu'il s'agit de la fiction de l'état instantané de la conscience au moment approprié, et à l'inverse une théorie des multiples façons selon laquelle le fait qu'un texte est fictionnel peut entrer en jeu (§3.1). Une approche de la dynamique sera proposée qui met l'accent sur la contrainte qu'impose la dynamique sur les états instantanés : un état instantané doit non seulement pouvoir représenter ou prendre en compte les aspects appropriés au moment particulier, mais il doit les représenter *d'une manière qui permet une pensée féconde de la dynamique*. En un mot, il faut les représenter de telle manière que la dynamique sera *continue* (§2.3). Dès que l'on caractérise les états instantanés de telle manière que leur dynamique est continue, de nouvelles conséquences suivront. Par exemple, on en vient comprendre certains aspects dynamiques de la lecture de la fiction, dont la mise en rapport du fait que le texte est fictionnel, les entrées et les sorties de l'immersion fictionnelles, les influences des attitudes ordinaires du lecteur sur la lecture, et, à l'inverse, les influences de la lecture sur ses attitudes ordinaires, en les rapportant aux propriétés spécifiques des états instantanés à des moments particuliers (Chapitre 3).

De l'autre côté, la conception du contexte proposé dans la théorie des contre-

factuels permet de rendre compte de phénomènes souvent négligés précédemment, dont par exemple la *pertinence* de l'antécédent par rapport au conséquent, le fait pour des sujets ou des aspects d'être *hors sujet* à un moment donné, ou encore les rapports entre les évaluations et les apparitions du même contrefactuel dans des contextes différents (§§5.2.1, et 5.2.2). Quant à la dynamique, comme dans le cas de la psychologie de la fiction, une conception du contexte approprié pour l'évaluation des contrefactuels est jugée également selon sa capacité à soutenir une pensée féconde de la dynamique des contextes, c'est-à-dire, une pensée de la dynamique comme continue : la notion de contexte proposé satisfera apparemment à cette exigence (§5.2.3). Enfin, la séparation des deux aspects d'une théorie de contrefactuels – le contexte instantané et les rapports entre ces contextes – implique une perspective différente sur la *logique*, en l'occurrence la logique des contrefactuels : les dites « règles logiques » pourraient, d'une part, caractériser les structures des contextes particuliers, ou autrement délimiter les transitions entre contextes qui sont « légitimes » ou « valides » dans un certain sens (§5.3.2 et Annexe A).

La question générale du rapport entre fiction et réalité, ou entre contrefactuels et réalité, est une question difficile, ne permettant pas de réponse complète facile. Cette thèse n'offre pas de réponse complète ; elle offre plutôt des cadres conceptuels pour penser la fiction et le contrefactuel dans leurs rapports à la réalité. Ces cadres permettent de voir ces rapports sous un nouveau jour, de rendre compte de certains phénomènes qui étaient jusque-là problématiques sinon carrément négligés, de comprendre certains subtilités des façons intéressantes et utiles, d'entamer une étude sur les aspects les plus épineux qui paraît prometteuse. Perspective féconde et intéressante, dissolution ou résolution de vieux problèmes, promesse : d'une enquête qui se veut méthodologique, vous ne pouvez guère exiger plus. La bonne réponse à une question ouverte n'est pas celle qui la clôt, mais celle qui montre la voie.

Première partie

Jouer à croire :

l'examen des processus
mentaux impliqués dans la
lecture des textes de fiction

Chapitre 1

Attitudes à l'égard de la fiction

1.1 La problématique de la fiction

1.1.1 Le seuil de la problématique : quatre questions

Quatre questions pour introduire la problématique.

A. Croyez-vous que Sherlock Holmes a existé ? Si vous répondez oui, on vous regardera avec curiosité et puis on vous expliquera que les romans de Sherlock Holmes ne sont que des fictions et que les personnages qui se trouvent dans des fictions n'ont pas nécessairement existé. D'autre part, si vous répondez non, on aura l'air surpris, et on vous demandera si à votre avis les histoires racontées au sujet de Sherlock Holmes sont le produit des hallucinations de Watson plutôt que de ses observations.

Il n'y a aucun doute qu'il y a ambiguïté ici, qu'il y a quelque chose qui manque, qui fait qu'on a envie de dire qu'on le croit et qu'on ne le croit pas. C'est cette chose-là qu'une analyse philosophique devrait tirer de l'ombre et mettre au clair.

En outre, ce genre d'ennuyeuse ambiguïté ne se limite pas au domaine cognitif : des effets semblables se retrouvent également dans le côté affectif de la fiction. A savoir, le fameux « paradoxe » de la fiction :

B. Éprouvez-vous des émotions envers des situations décrites dans les fictions ? Une réponse négative passe sous silence les sentiments que vous ressentez lors de la lecture ou de la considération d'un texte de fiction. D'un autre côté, une

réponse positive se heurterait à votre savoir que les situations qui y sont décrites n'ont pas vraiment lieu : car il semble qu'il faille croire à une situation, fût-ce partiellement, pour éprouver une émotion à son égard. En outre, votre réaction à la situation décrite n'est pas celle de quelqu'un qui éprouve normalement une émotion envers une telle situation : vous n'agissez pas pour sauver l'héroïne, pour empêcher un désastre imminent, ou pour sanctionner le comportement déplorable d'un personnage.

De toute évidence, cette équivoque sur le plan affectif reflète celle qui a été précédemment repérée sur le plan cognitif.

Les difficultés que pose la fiction ne sont pas limitées cependant à de simples ambiguïtés des attitudes ou des états mentaux à son égard. Plus épineuse peut-être est la question des inférences que l'on fait à partir des énoncés du texte fictionnel, lesquelles peuvent mettre en jeu des faits réels non explicités dans le texte.

C. Supposez que vous lisez un texte de fiction dont le héros (supposé humain) est dans une position où il risque de tomber d'un escarpement élevé. Vous en concluez sans doute qu'il est en position de danger. Ce faisant, vous devez vous appuyer sur des énoncés vrais en réalité mais qui ne sont pas explicités dans le texte : il vous faut admettre par exemple que les humains ne survivent pas à de telles chutes, alors qu'à strictement parler cet énoncé n'appartient pas au texte. En faisant entrer les vérités réelles, vous risquez de tomber dans la confusion : le fait que le héros n'existe pas est une vérité réelle que vous connaissez bien ; de l'autre côté, vous devriez croire également que le héros existe, dans la mesure où ceci est une condition pour qu'il puisse être en danger. Comment évitez-vous cette conclusion ? Plus particulièrement, si vous faites un tri entre les faits de la réalité que vous mettez à l'œuvre dans votre lecture d'un texte de fiction et ceux que vous ne vous permettez pas, où situez-vous la ligne de partage ?

Le travail philosophique appelé ici ne paraît plus trivial : en se débarrassant d'une contradiction menaçante et gênante, on risque de déboucher sur une question encore plus difficile et complexe.

Enfin, il y a un autre versant de cette question, concernant non pas les faits (connus) réels que vous mettez en œuvre dans votre lecture de la fiction, mais plutôt les acquis de cette lecture qui s'infiltreront dans votre conception du réel.

D. Vous avez rencontré le personnage de Sherlock Holmes dans les textes fictionnels appropriés, et certainement pas dans la réalité. Or, vos attitudes à l'égard ce personnage semblent parfois avoir une teneur dans la réalité. Par exemple,

vous croyez que Sherlock Holmes est un personnage fictionnel, mais le statut de cette croyance n'est pas clair : c'est une croyance relative à quelque chose que l'on croit ne pas exister, ce qui ne serait que difficilement une croyance, encore plus une croyance cohérente. En outre, la figure de Holmes pourrait vous servir pour incarner ou pour exemplifier certaines caractéristiques, telles l'intelligence ou l'ingéniosité ; vous pourriez croire que quelqu'un (de réel) est « un Holmes ». Quels sont les statuts de ces croyances ? Ou encore, la lecture des *Aventures de Sherlock Holmes* pourrait bien avoir changé votre avis sur l'état *réel* de l'existence, sur le mal et sur la nature de l'intelligence par exemple. Quels sont les effets de la lecture d'un texte fictionnel sur votre conception de la réalité ? Comment comprendre la tension apparente entre ces éventuels effets « réels » de textes fictionnels et votre connaissance de leur fausseté ?

Ces quatre questions appellent apparemment des réponses philosophiques. Dans les chapitres suivants, il s'agit de proposer une théorie des attitudes du lecteur à l'égard de la fiction qui offre des réponses à ces questions. Dans cette mesure, ces questions résument la problématique de la première partie de la thèse. Avant de développer cette théorie, il convient de préciser les difficultés et les défis qui se posent et considérer les approches qui en ont été déjà proposées. Le chapitre présent est consacré à cette tâche.

1.1.2 Le contour de la problématique

Où l'on arrive à mieux cerner la question générale dont s'occupera ce travail comme celle du rapport entre la fiction et la réalité, en tant qu'elles apparaissent dans la psychologie du lecteur. Pour étudier ce rapport, on adopte une position « internaliste », permettant seulement des concepts et des discussions qui s'appliquent au plan du mental du lecteur, et non pas, par exemple, au plan de l'auteur, des actes communicationnels ou de la réalité en quelconque sens ontologique ou métaphysique plus fort. Enfin, il s'agit, par hypothèse, du lecteur qui a connaissance du fait que le texte qu'il lit est fictionnel.

Les quatre questions posées dans le paragraphe précédent ont à affaire un sujet commun : *le rapport entre la fiction et la réalité*. La question **A.** concerne le rapport entre les attitudes cognitives à l'égard du réel et celles à l'égard de la fiction. Pour un énoncé comme « Sherlock Holmes a existé », il semble y avoir *un sens* où l'on peut dire le croire – le sens d'une croyance « fictionnelle » – et *un autre sens* où l'on ne le croit pas – le sens de la croyance réelle ou usuelle.

La question est celle de ces deux sens et du rapport entre eux. La question **B.** a trait aux rapports entre d'une part les « croyances *fictionnelles* » à l'égard des situations décrites dans les textes de fiction et d'autre part les sentiments sinon les émotions *réelles* que l'on éprouve à leur égard. Les questions **C.** et **D.** portent respectivement sur « l'intrusion » de la réalité dans la fiction, si l'on puis dire, et inversement sur le « débordement » du fictionnel sur la réalité.

En effet, ces quatre questions ne sont que des facettes différentes de la question générale à laquelle est consacrée cette partie de la thèse, à savoir celle du *rapport entre la fiction et la réalité* en tant qu'elle se pose dans le domaine spécifique de la *psychologie du lecteur*. Dans ce chapitre et les deux chapitres suivants, il s'agit des états mentaux d'un lecteur *typique* ou *ordinaire* à l'égard de la fiction, et du rapport avec ses états « usuels » à l'égard de la réalité. Les états mentaux en question seront *non seulement* ceux qui font occurrence *pendant* la lecture de la fiction, *mais également* les états mentaux à *des moments postérieurs* à la lecture qui se rapportent d'une manière ou d'une autre au texte de fiction.

En restreignant le questionnement à la psychologie du lecteur et à elle seule, on écarte un certain nombre d'autres questions, relatives par exemple à l'écriture de la fiction, à l'acte communicationnel fictionnel, à l'existence ou à la nature des entités fictionnelles ou des mondes fictionnels. Il se pourrait que les questions posées, l'approche empruntée et les réponses recommandées ici aient des analogues ou des conséquences relatives à ces autres questions ; pourtant, une considération spécifique de ces autres questions déborderait les possibilités de cet essai.

En outre, la perspective de ce questionnement fait fi d'un certain nombre de questions « ontologiques », « métaphysiques » et « sémantiques » à l'égard de la fiction. Il *ne sera pas question* de l'existence des entités que le lecteur imagine ou croit exister, ni de la vérité réelle des énoncés qu'il considère, ni de la manière dont il en vient à apprendre des vérités de la réalité ; il s'agira seulement de la différence qu'il *conçoit lui-même* entre ce qui existe et ce qui n'existe pas, entre ce qui est vrai et ce qui ne l'est pas, entre ce qu'il a appris de la réalité et ce qu'il a appris du texte de fiction. À l'intérieur de ces limites, le questionnement est entièrement général : il s'agira et bien des attitudes du lecteur envers le texte et bien de « l'univers fictionnel » qu'il « se figure dans son imagination » et bien de ses comportements pertinents pour le texte (telles ses affirmations) en tant qu'il les saisit. Si, par souci de commodité, on travaille majoritairement avec les attitudes du lecteur envers les énoncés, la discussion possède effectivement une portée plus grande et plus générale.

Or, cette perspective – « internaliste » dans la mesure où elle s’occupe uniquement du point de vue du lecteur, de ce qu’il éprouve et de la conception qu’il se fait de la fiction et de sa lecture¹ – n’est pas aussi étroite que l’on pourrait la penser.

Tout d’abord, l’adoption de cette perspective n’empêche nullement le théoricien de prendre son propre point de vue, à la troisième personne pour ainsi dire, sur le mental du lecteur, ni d’employer ses propres concepts et de poser ses propres questions. Elle ne fait qu’imposer une sorte d’économie méthodologie, dans la mesure où elle limite les questions qui pourraient être posées, et surtout les concepts qui pourraient être employés, à celles et à ceux qui ont un sens dans la perspective « interne ». Donc, par exemple, dans cette perspective il ne pourrait pas y avoir de discussion des entités, attitudes et actions du monde externe au lecteur qui « réalisent » ou correspondent aux phénomènes décrits dans le texte de fiction ou ayant trait à sa lecture. Par conséquent, cette perspective interdit l’analyse des phénomènes, des effets et des modalités mentales ayant trait à la lecture d’un texte de fiction en s’appuyant sur une quelconque supposition relative à son statut ontologique, épistémique ou existentiel.

En outre, il n’en découle aucune impossibilité de rendre compte des *influences* sur la lecture des aspects externes au texte de fiction particulier, telles les influences « inconscientes » intertextuelles, sociales ou même des suggestions linguistiques. Car, pour que ces influences « externes » s’exercent sur le lecteur, sa lecture et son appréciation du texte, il faut qu’elles soient déjà « entrées » dans la psychologie d’une manière ou d’une autre. Par exemple, pour que les références intertextuelles puissent influencer sur sa lecture, il faut que le lecteur ait déjà lu les textes appropriés, ou du moins qu’il en ait pris autrement connaissance. La perspective « internaliste » n’empêche en rien d’aborder les questions de telles influences « externes » ; elle exige seulement le recours aux moments où et aux manières dont leurs supports sont entrés dans la vie mentale du lecteur et leurs différentes conditions d’entrée. Comme l’a fait remarquer Wittgenstein, le solipsisme au sens strict coïncide avec le réalisme².

¹Le fait de cibler la fiction en tant que conçue et vécue de l’intérieur n’implique aucun *rejet* d’une réalité hors la psychologie du lecteur, mais seulement une *mise entre parenthèses* de cette réalité pour les besoins de ce travail. Il ne s’agit donc pas du « solipsisme méthodologique » de Fodor (« les processus mentaux ... n’ont aucun accès aux propriétés sémantiques des représentations », 1980, p65) ou de Putnam (« aucun état psychologique ne présuppose l’existence de quelconque individu autre que le sujet de l’état », 1975, p220) en ceci qu’il ne s’agit pas de l’accès ou de la présupposition de la réalité. Peut-être l’attitude la plus proche dans la littérature est-elle l’*epochè* husserlienne (1913a, §32), bien que, à la différence d’Husserl, le questionnement adopté ici ne soit nullement eidétique.

²Wittgenstein (1922, §5.64).

Pourtant, la question du rapport entre la fiction et la réalité ne peut être formulée dans une perspective « internaliste » qu'à condition que l'on suppose les deux termes de ce rapport, à savoir, que l'on suppose qu'il y ait « dans le mental » du lecteur la distinction entre le fictionnel et le réel. Pour permettre l'étude des états et des phénomènes mentaux ayant trait à la fiction spécifiquement (plutôt qu'au leurre par exemple), on admet l'hypothèse de travail suivante :

Le lecteur du texte de fiction sait, avant même de se mettre à sa lecture, que le texte est fictionnel.

Sans la connaissance de la part du lecteur du fait que le texte est fictionnel, il n'y a pas de distinction entre fiction et réalité *dans sa psychologie* (relative à ce texte) à étudier. Si bien que la question générale posée ici pourrait être vue comme celle des *conséquences* de cette connaissance.

Ayant précisé les aspects de base du questionnement qui dominera cette partie, il convient de l'approfondir quelque peu. En remarquant deux aspects du rapport entre fiction et réalité qui sont mis en jeu par les quatre questions de §1.1.1, on viendra non seulement à mieux apercevoir l'enjeu, mais on débouchera naturellement sur un genre de réponse répandue qui prend comme point de départ exactement ces deux aspects.

1.1.3 Le cœur de la problématique

Où deux aspects importants et généraux du rapport entre fiction et réalité sont repérés : une distinction stricte entre fiction et réalité, mais aussi une certaine similitude entre elles. La tension entre ces deux aspects sous-tend, dans des formes différentes, les quatre questions posées au début. Il s'agit, dans cette partie de la thèse, de comprendre cette tension dans sa complexité.

La structure générale du rapport entre fiction et réalité est tellement bien connue qu'elle en est presque banale. Elle comprend les deux aspects suivants :

- Une certaine *similitude* entre fiction et réalité ;
- Une *distinction* cruciale entre elles.

Exprimé de cette façon grossière, il paraîtrait à première vue qu'il y ait une tension entre ces deux aspects. On est certainement enclin à penser que plus la similitude entre fiction et réalité s'accroît, plus la distinction entre elles s'affaiblit. Les quatre questions posées dans §1.1.1 mettent en jeu, chacune à sa manière propre, ces deux aspects et la tension entre eux.

D'une part, il y a (au moins) trois manières de lire les questions **A.** et **B.**,

et par conséquent trois questions différemment spécifiées qui peuvent en être tirées.

Premièrement, on pourrait voir dans la question **A.** une question relative aux entités fictionnelles en tant que le lecteur se les figure : elles s'apparentent « dans la vue de l'esprit » aux « entités réelles ». Cette question est le versant psychologique de la fameuse question des « entités fictionnelles » – qu'est-ce que Sherlock Holmes, personnage dont on connaît apparemment beaucoup et auquel on fait apparemment référence, mais qui manque des propriétés logiques usuelles, telle la « complétude » (il n'est ni vrai ni faux qu'il a 1890 cheveux sur la tête, le 14 juin, 1887) ?³

Selon une deuxième lecture de la question **A.**, elle a trait à ce que Schaeffer appelle la « compétence fictionnelle »⁴, laquelle comporte aussi bien la possibilité pour le lecteur de « s'oublier » dans sa lecture – de s'engager dans ce que Schaeffer appelle « l'immersion fictionnelle »⁵ –, que sa capacité de « se rappeler » la réalité d'alentour en cas de nécessité. La question de l'existence de Holmes, posée dans une situation « ordinaire », représente un éventuel écueil pour le lecteur (s'il ne reconnaît pas la distinction entre fiction et réalité, il répondra à tort que Holmes existe) : en évitant de confondre la fiction et la réalité, il fait épreuve de sa « compétence fictionnelle ». À part les conséquences de la distinction que fait le lecteur entre fiction et réalité, ses comportements à l'égard du texte fictionnel ressemblent à ses comportements face à un texte réel.

Selon une troisième lecture, les questions **A.** et **B.** concernent, dans les domaines cognitif et affectif respectivement, la nature de la distinction et de la ressemblance entre les attitudes impliquées dans la lecture et la compréhension de la fiction et les attitudes usuelles. Du côté cognitif, la croyance ordinaire « s'insère » dans une structure cognitive, où elle est en relation avec d'autres croyances et attitudes, telle le désir. Les « attitudes » cognitives relatives aux événements et aux personnages décrits dans le texte de fiction – les croyances « fictionnelles », pour ainsi dire – s'insèrent, pour leur part, dans une structure autre mais tout à fait semblable. De même que les croyances « usuelles », les croyances « fictionnelles » entretiennent des relations importantes les unes avec les autres : il n'est apparemment pas possible d'avoir la croyance « fictionnelle » que Holmes a habité à Baker Street sans aussi en avoir la croyance « fictionnelle » que Holmes a existé par exemple ; ou encore, la croyance « fictionnelle » que Holmes a habité à Baker Street semble entraîner la croyance « fictionnelle »

³Cette question est discutée par Parsons (1975), Walton (1990, Ch. 10), Currie (1990, Ch. 4), parmi d'autres.

⁴Schaeffer (1999, Ch. III, §3).

⁵Schaeffer, *op cit.*, Ch. III, §4.

qu'il a dû payer des impôts britanniques. Et de même que la structure cognitive ordinaire est largement « cohérente » et « justifiée », au sens où les contradictions tendent à être « dissoutes » et les croyances tendent à y entrer sur la base de leurs rapports à d'autres croyances et aux perceptions ; de même la structure des croyances « fictionnelles » est largement « cohérente » et « justifiée », en l'occurrence par rapport au texte de fiction. Du côté affectif, tout comme les émotions se rapportent à la structure cognitive, dans la mesure où elles dépendent des croyances et influent sur les actions⁶, les affects relatifs à un texte de fiction sont en relation avec les croyances « fictionnelles » à son égard : ils s'expliquent et ils se justifient en grande partie par leur rapport à des attitudes cognitives envers le texte, ils semblent exercer une influence sur des attitudes postérieures, et ainsi de suite⁷. Il y a donc des ressemblances structurelles avec les attitudes et les états mentaux ordinaires, mais également des différences importantes : ni les croyances « fictionnelles » ni les affects éprouvés pendant la lecture ne mènent aux actions qui se produiraient en cas des croyances ou des affects usuels correspondants⁸.

D'autre part, dans la mesure où elles s'occupent des points où la distinction entre fiction et réalité s'efface, où la réalité passe dans la fiction et la fiction s'infiltre dans la réalité, les questions **C.** et **D.** concernent précisément l'extension de la distinction cruciale entre fiction et réalité. Il s'agit non pas des différences et des similitudes entre les « genres » d'attitudes, de comportements ou d'entités « fictionnelles » et « réelles », mais plutôt des chevauchements entre la fiction et la réalité. Non pas des questions : *que* sont les attitudes, les entités ou les comportements impliqués dans la fiction ? Mais plutôt des questions : *quelles* entités ou attitudes « réelles » sont impliquées dans la lecture de la fiction, et *quels* rôles jouent-elles ? *Quelles* entités ou attitudes « fictionnelles » acquièrent des statuts « réels », et *lesquels* ? Dans le sens où ces deux questions portent sur le chevauchement ou la relation difficile, complexe et entrecroisée entre fiction et réalité, toutes les deux ont affaire à une certaine similitude entre fiction et réalité. En revanche, elles diffèrent dans le *sens* de ce chevauchement ou de cette similitude. **C.** s'occupe des aspects de la lecture qui empruntent à la réalité – d'une similitude de la fiction à l'instar de la réalité. Par contraste, **D.** a affaire aux possibles emprunts à la fiction – une similitude où la réalité suit la fiction.

Ces deux questions ont été posées, sous plusieurs formes, dans la littérature.

⁶Pour une discussion plus complète, voir §1.3.4.

⁷Sur ce point, voir Currie (1990, §§5.3 et 5.4) par exemple.

⁸Comme on aura l'occasion de discuter longuement, les philosophes tels Walton, *op cit.*, Ch. 1 et 7, et Currie, *op cit.*, §1.5 et Ch. 5, ont abordé cette question.

D'une part, la question **C.**, relative aux apports de la réalité à la lecture, est fortement reliée à la question de la dite « *concrétisation* » des entités ou des événements fictionnels, aussi bien, plus généralement, qu'à la question des « *vérités fictionnelles* ». La notion de « *concrétisation* », introduite par Ingarden⁹, dénote le phénomène par lequel le lecteur « étoffe » les propos du texte – la description d'un personnage par exemple – pour s'en faire une image plus « complète » ou « concrète ». Une « *vérité fictionnelle* » est un énoncé qui est autorisée par un texte de fiction ; cette notion a été soumise à une analyse célèbre par Lewis¹⁰, sur laquelle on reviendra dans la seconde partie de ce travail. Dans les deux cas, il s'agit d'un apport de la réalité : en étoffant les propos du texte dans une concrétisation, ce sont les croyances, préjugés, inclinations et avis *ordinaires* du lecteur qui sont apparemment mis en jeu ; en « inférant » les vérités fictionnelles vérité fictionnelle, comme l'a montré l'exemple précédent du héros en danger (§1.1.1), ce sont les croyances ou connaissances *réelles* qui entrent en jeu. En posant la question générale **C.**, on pose donc ces deux questions.

D'autre part, la question **D.**, ayant trait aux apparitions ou effets de la fiction sur le terrain de la réalité, a affaire aux usages des entités ou des événements fictionnels dans des contextes « réels », aussi bien, plus généralement, qu'aux « effets » ou aux « influences » de la lecture de la fiction. En premier lieu, il s'agit des énoncés figurant des termes « fictionnels » et donc non dénotants (Sherlock Holmes) qui ne sont pas vrais *dans* la fiction mais plutôt vrais *de* la fiction (« Sherlock Holmes est un personnage fictionnel »), énoncés dont l'analyse est d'ailleurs une question débattue par de nombreux philosophes, dont Walton et Currie¹¹. Ces énoncés sont employés pour l'essentiel dans des contextes « réels » pour discuter des entités ou des événements *fictionnels*. En second lieu, il est question des « rapprochements » entre entités fictionnelles et entités réelles, ou plus précisément dit des « prédictions » des entités réelles par des termes provenant de la fiction (« Untel est un 'Don Quichotte' »), lesquels sont discutés notamment par Goodman¹². Dans ces cas, le terme 'Don Quichotte', terme introduit et développé dans une œuvre de *fiction*, est employé dans la *réalité*. En dernier lieu, il y a la question générale des éventuels « effets » de la lecture de la fiction sur la conception qu'a le lecteur de la réalité. Iser, par exemple, met l'accent sur le fait que certains textes littéraires (et fictionnels)

⁹Ingarden, 1965, p214 : « le lecteur, lors de la lecture et de la saisie esthétique de l'œuvre, dépasse le donné textuel et *complète* à nombreux égards les « objets » figurés. » Voir également, Ch. 13 (notamment note 9, p 285), et Ingarden (1968, §11).

¹⁰Lewis (1978).

¹¹Walton, *op cit.*, Ch 10 & 11, Currie *op cit.*, §4.11.

¹²Goodman (1978, §VI.5).

contestent, défendent, mettent en question ou renforcent certains « systèmes de pensées » ; de façon similaire, Ricoeur met l'accent sur l'acte de « re-figuration du monde » (en l'occurrence d'un monde temporel) impliqué dans la lecture¹³. Ces éventuels « effets » de la lecture de la fiction sur la conception de la réalité dépassent la simple distinction entre fiction et réalité, au profit d'une relation plus complexe. Il importe de souligner que, dans la perspective de ce travail, ces « effets » ne représentent que des *possibilités* des textes fictionnels et aucunement des *nécessités* : comme l'a été remarqué dans l'Introduction, il s'agit ici de la fiction en général, c'est-à-dire aussi bien des textes qui cherchent et qui réussissent à promouvoir certains effets, que des textes qui ne les cherchent ni ne les produisent.

La question du rapport entre fiction et réalité, en tant qu'elle se pose à travers les quatre questions de §1.1.1 (et dans la sphère du mental du lecteur), se précise de manière suivante. C'est la question de la tension entre la similitude entre fiction et réalité et la distinction entre elles : similitude et distinction dans les attitudes, les entités et les comportements, similitude et distinction instaurées par les empiètements des entités, vérités ou thèmes réels dans la fiction et fictionnels dans le réel. L'enjeu est de *comprendre* toute cette tension, non pas de l'étouffer ni de la dissoudre. Autrement dit, il s'agit de comprendre le sens dans lequel elles se ressemblent et le sens dans lequel elles se distinguent. La question du rapport entre fiction et réalité est celle de l'*équilibre* entre leur similitude et leur différence. Ce qui est tout à fait naturel : dire qu'il y a ressemblance ou distinction n'est rien dire si on ne précise pas *quelles* ressemblances et *quelles* distinctions.

1.2 Approches générales de la fiction

Malgré les nombreuses différences qui opposent les théories qui s'occupent, d'une manière ou d'une autre, de la fiction, il s'avère, comme on le montrera dans le premier paragraphe de cette section, que la grande majorité souscrivent à une stratégie commune. Cette stratégie, que l'on appelle la stratégie « mimétique », consiste à poser d'abord la distinction entre un côté « fiction » et un côté « réalité » pour ensuite aborder la question du rapport entre eux. Or, comme on montrera en profondeur dans §§1.4 et 1.5, cette stratégie ne fournit pas de réponse adéquate aux questions relatives à la fiction et à son rapport à la réalité qui ont été formulées dans la section précédente.

Il s'agira donc de développer une nouvelle théorie, qui emprunte une autre

¹³Iser (1976, par exemple Ch. 3), Ricoeur (1983, notamment 1^{re}Partie, Ch. 3).

stratégie. Dans cette section, on introduit et motive brièvement l'idée de base qui sous-tend cette stratégie. Elle s'apercevra de manière particulièrement claire par le biais d'un détour par la notion de jeu, lequel comporte une parenté reconnue avec la fiction. Le jeu aussi bien que la fiction entre dans deux sortes de relation avec la réalité. D'une part, il s'oppose à la partie de la réalité qui est « annulée » ou « mise en suspension » par les règles du jeu ; d'autre part, il s'appuie sur une partie de la réalité qui est toujours « en vigueur » dans le jeu (§1.2.2). La stratégie « mimétique » prend comme point de départ l'opposition entre la fiction et réalité ; par contraste, la stratégie qui sera employée dans ce travail commencera à un niveau commun à la réalité et la fiction. À ce niveau, la distinction entre les deux n'est pas donnée au départ. Plutôt, il s'agira de *retrouver* cette distinction, et par conséquent les détails du rapport entre fiction et réalité, en travaillant à partir du niveau qui leur est commun. Dans §1.2.3, on esquissera les grandes lignes de cette stratégie et de sa mise en œuvre dans le cas de la psychologie de la lecture.

1.2.1 La stratégie « mimétique »

Où l'on considère les deux grandes approches souvent prises à l'égard de la fiction. Ce qui importe ici n'est pas leur différence mais leur communauté : elles emploient les deux la dite stratégie « mimétique », consistant à poser initialement la distinction entre fiction et réalité pour ensuite venir au rapport entre elles, lequel est invariablement de l'ordre de la similitude.

Il est coutume de distinguer, pour ce qui concerne les approches proposées de la fiction, deux traditions : l'une, qui remonte jusqu'à Platon, souligne l'aspect *imitatif* de la fiction, en pensant par exemple l'acte d'écrire, le texte fictionnel lui-même, les propos du texte, sa communication, ou encore sa lecture comme des « imitations » des actes, textes ou faits réels ; l'autre, qui remonte aux romantiques du XVIII^e (et dont certains ont cru repérer des traces dans la *Poétique* d'Aristote) se propose de penser l'art en général et la fiction en particulier comme une activité *créative*. En bref, la première aborde la fiction comme du *mimésis* au sens de l'*imitation*, alors que la dernière privilégie un autre sens du mot *mimésis* qui a été récemment mis en valeur, celui de la représentation¹⁴ ou mieux de la « mise en représentation »¹⁵. Or, si l'on prend des choses à un niveau suffisamment général, on a affaire à deux exemples d'un seul et même schème

¹⁴Voir la discussion pp24-30 de l'introduction de Magnien (Aristote, 1990), aussi bien que la traduction de Dupont-Roc et Lallot (Aristote, 1980).

¹⁵Ricœur (1983, p69).

d'approche de la fiction, une seule stratégie, que l'on baptise, en jouant sur le double sens du mot, la *stratégie « mimétique »*. Cette stratégie rend compte des deux aspects du rapport entre fiction et réalité – leur *distinction* et leur *similitude* (§1.1.3) – en posant ou supposant *d'abord* les deux termes *différents* – la fiction et la réalité – pour *ensuite* considérer le *rapport* entre ces termes. Dans ce sens, elle commence avec la distinction entre la fiction et la réalité, pour en venir après, d'une manière ou d'une autre, à leur similitude. Pourtant, il y a des raisons pour douter de l'efficacité de cette stratégie. Dans les sections ultérieures de ce chapitre, on examinera les applications de cette stratégie à la question psychologique de la fiction, en mettant en évidence certaines faiblesses qui leurs touchent.

Un représentant contemporain typique de l'approche *imitative* de la fiction, c'est-à-dire de l'approche qui traite la fiction comme une « imitation » ou une « feinte » de la réalité, ou plus généralement comme « modélisée » d'après elle, est la théorie de « make-believe » défendue par Walton et Currie. Selon Walton, la fiction consiste en un jeu, joué par le lecteur avec le texte, qu'il appelle le jeu de « make-believe »¹⁶. Dans ce jeu, un lecteur « imagine » (pour employer le terme de Walton) ou « make-believe » (pour employer le terme de Currie) les énoncés du texte. Ce « make-believe » n'est pas le *faire croire* (notamment elle ne signifie nullement que l'auteur *fait que* le lecteur *croit* les énoncés du texte) ; en effet, Walton et Currie veulent désigner plutôt une certaine *attitude mentale* du lecteur, qu'une force externe exercée sur lui par l'auteur. Le « make-believe » est ainsi de l'ordre du « *faire comme si* », ce qui entraîne la connotation de feinte présente dans le terme anglais. En plus, à côté de cette *attitude* de *faire comme si*, il y a un *opérateur* de *faire comme si* : « il est « *faire comme si* » que *X* » (*it is make-believe that X*) ou plus correctement « il est vrai dans le jeu de « *faire comme si* » que *X* » si *X* est à imaginer ou à *faire comme si* (au sens de l'attitude) dans le jeu de *faire comme si*¹⁷. Pour Walton surtout, un énoncé est une « vérité fictionnelle » s'il est du *faire comme si* (au sens de l'opérateur)¹⁸. De fait, l'opérateur de *faire comme si* « imite » la vérité, l'attitude de *faire comme si* « imite » la croyance et le rapport entre eux « imite » le rapport entre vérité et croyance : « Imaginer vise à ce qui est fictionnel, comme croire vise à ce qui est vrai. Ce qui est vrai est à croire ; ce qui est fictionnel est à imaginer. »¹⁹

¹⁶Walton (1990, Ch. 1).

¹⁷Walton, *op cit.*, p39 : « les propositions fictionnelles sont des propositions qui sont à imaginer [dans le jeu de « make-believe »]. » Voir également Currie (1990, §§1.5 et 2.4).

¹⁸Walton, *op cit.*, p35. Currie souligne par ailleurs la différence entre vérité fictionnelle et ce qui est du *faire comme si* dans un jeu de *faire comme si* dans 1990, §2.4.

¹⁹Walton, *op cit.*, p41.

De l'autre côté, l'approche créative de la fiction, développée par des théoriciens tels Goodman, Ricoeur, et Iser, se concentrent sur les possibilités « créatives » de la fiction, ses possibilités de « re-figurer », de « re-former », d'« influencer », de « contrebalancer » la (conception de la) « réalité ». Comme le dit Goodman, « Cervantes, Bosch et Goya, autant que Boswell, Newton et Darwin, prennent et défont, font et reprennent des mondes familiers, en les refaçonant de des manière remarquable ... »²⁰ Pour sa part, Ricoeur parle de la « configuration nouvelle » produite dans ou par la lecture²¹. Enfin, Iser ne cesse de souligner la particularité du « texte littéraire » (dont tous ses exemples sont des textes de fiction) : « La relation unique entre le texte littéraire et la 'réalité', sous forme des systèmes de pensée ou modèles de réalité, tient à ceci ... [le texte] représente une réaction contre les systèmes de pensées qu'il a choisis »²².

Ces deux traditions ont ceci de commun : elles posent les deux une *distinction* entre un côté « fiction » et côté « réalité », pour ensuite penser à leur *rapport* ou leur *similitude*. D'un côté, dans la théorie de Walton et Currie, l'activité propre à la fiction se distingue de l'activité ordinaire par le fait qu'il s'agit d'un jeu, or elle « ressemble à » (ou « imite ») l'activité réelle dans la mesure où elle est du *faire comme si*. Les énoncés relatifs à la fiction se distinguent des énoncés ordinaires par le préfixe « il est vrai dans le jeu de *faire comme si* que ... », or la vérité dans le jeu, par le fait même qu'il s'agit du *faire comme si*, ressemble à la vérité au sens ordinaire²³. Enfin, l'attitude du lecteur relative à la fiction (l'imagination au sens de Walton, le *faire comme si* au sens de Currie) est différente de son attitude relative aux faits réels (la croyance)²⁴, or les deux auteurs s'accordent à ce qu'elles se ressemblent, dans la mesure où les *faire comme si* prennent place dans une structure analogue à la structure cognitive où s'étalent les croyances. Par exemple, de même qu'il y a un certain « idéal de cohérence » et de « justification » des croyances, les *faire comme si* sont, autant que possible, cohérents et justifiés ... par rapport au texte de fiction²⁵. En somme, en ramenant la fiction au « *faire comme si* », on la *distingue* de la

²⁰ Goodman (1978, p104-5).

²¹ Ricoeur (1983, p12). Ricoeur concentre en l'occurrence sur la question de la « refiguration » du temps.

²² Iser (1976, p72).

²³ « La fictionalité s'avère en effet être de plusieurs manières analogue à la vérité », Walton, *ibid.*

²⁴ « Ce qui distingue la lecture d'un texte de fiction de la lecture d'un texte non fictionnel est ... l'attitude relative au contenu de ce que nous lisons : *faire comme si* dans un cas, croyance dans l'autre. » Currie, *op cit*, p 21.

²⁵ Walton, §1.5. Currie (p205 *sq.*) pousse encore plus loin la ressemblance en proposant les « make-desires » qui entrent dans les mêmes rapports avec les « make-beliefs » que les désirs (usuels) avec les croyances (usuelles).

réalité, mais en même temps on la considère comme *similaire* à la réalité, sur le mode *imitatif*. De l'autre côté, Goodman, Ricoeur et Iser commencent par poser la question du rapport entre deux choses prises pour distinctes – la fiction et la réalité (en tant qu'elle est conçue) – pour ensuite conclure que la première « travaille », « re-forme » ou « influence » la dernière. Du reste, dans le sens où la re-figuration du « monde » (réel) opérée par la fiction a pour conséquence une certaine parenté entre ces deux termes, on peut dire que ces auteurs retrouvent une sorte de similitude après avoir posé leur distinction.

L'attention mise ici sur la communauté entre les deux traditions n'implique cependant aucun rejet du fameux contraste entre elles : si elles admettent les deux de partir de la distinction entre fiction et réalité pour retrouver ensuite leur ressemblance, il ne reste pas moins qu'elles s'opposent sur le *sens* de cette ressemblance. La tradition imitative traite la réalité conçue comme pré-donnée, de telle sorte que toute similitude entre fiction et réalité est conçue comme une imitation *par* la fiction *de* la réalité. L'être fictionnel et l'attitude de *faire comme si*, par exemple, sont conçues *sur le modèle* de la vérité et la croyance bien connues. Par contraste, la tradition créative met en avant la possibilité pour la fiction de déformer ou re-figurer la réalité pré-donnée, et donc la possibilité pour la réalité (ou mieux, la conception de la réalité) de « suivre » ou « se régler » sur la fiction dans certains cas, de telle sorte que c'est la réalité qui vient ressembler à la fiction. Si, comme le soutient Iser, la marque de la fiction est « le franchissement de frontières »²⁶, si la fiction prend un « répertoire » des conceptions et des systèmes de pensées de la réalité et les réorganise²⁷, alors une similitude entre réalité et fiction (selon cette dimension) n'est rétablie que si la réalité « rattrape » la fiction qui « l'a dépassée ». La similitude est *a priori* un rapport symétrique (X est semblable à Y si et seulement si Y est semblable à X) ; néanmoins, la tradition imitative accorde une priorité à la réalité, et par conséquent valorise le sens où la fiction ressemble à la réalité, alors que la tradition créative accorde une priorité à la fiction, et par conséquent au sens où la réalité est ou devient semblable à la fiction. Cette différence est sans doute liée aux différences des intérêts et des buts des théories appartenant à chaque tradition ; c'est cependant leur communauté plutôt que leur différence qui importe ici.

En conséquence, il est légitime de ranger ces deux genres d'approche sous le signe d'une seule stratégie pour aborder la fiction, que l'on appelle la *stratégie « mimétique »*. Elle consiste à poser d'abord la distinction entre fiction et réalité,

²⁶Iser (1991, Ch. 3).

²⁷Iser (1976, Chs. 3 et 4).

pour ensuite retrouver leur rapport, voire leur similitude. On se chargera dans les sections suivantes (§§1.4 et 1.5) de montrer que cette stratégie « mimétique », en tant qu'elle a été appliquée de diverses façons à la question psychologique de la fiction, n'est pas la bonne. En un mot, la difficulté tient au fait que la stratégie ne permet guère de spécifier ni de comprendre les *détails* du rapport entre la réalité et la fiction. L'application de la stratégie « mimétique » à la question *sémantique*, c'est-à-dire la question des « énoncés fictionnels » tels « Holmes est un détective », permet d'illustrer ce genre de difficulté. Souvent, on pose *initialement* une distinction entre énoncés relatifs à la fiction et les énoncés « ordinaires » sous forme d'un opérateur comme « dans la fiction f, ... »²⁸ : à la différence des énoncés « ordinaires », les énoncés ayant trait à la fiction (« Holmes est un détective ») sont supposément des ellipses des énoncés préfixés par cet opérateur (« dans les Aventures de Sherlock Holmes, Holmes est un détective »). *Ensuite*, on analyse cet opérateur, tantôt en mettant en œuvre des techniques empruntées à la sémantique des énoncés ordinaires²⁹ et donc une similitude avec l'analyse des énoncés ordinaires, tantôt en s'appuyant explicitement sur la notion de « similitude » avec la réalité³⁰. On n'est pourtant pas parvenu à comprendre parfaitement les détails de cette similitude avec la réalité ou avec l'analyse des énoncés ordinaires. Par exemple, Lewis analyse l'opérateur fictionnel en termes de contrefactuels (grossièrement, des conditionnels subjonctifs « si X était le cas, alors Y serait le cas »), qui, à leur tour, sont analysés en termes de *similitude* entre mondes possibles (*grosso modo*, il s'agit des mondes possibles les plus semblables au monde réel où les énoncés du texte sont énoncés comme vérités)³¹. Cette similitude entre mondes, à laquelle sont renvoyés les contrefactuels ainsi que la vérité dans la fiction, n'est pas très bien comprise, variant, comme l'admet Lewis lui-même, en fonction de « la pratique et le contexte »³². Par conséquent, la question *détaillée* du rapport entre fiction et réalité, qui se réduit selon cette analyse à la question *détaillée* de la relation de similitude entre mondes possibles, ne trouve pas de réponse satisfaisante. En reconnaissant le rapport entre fiction et réalité comme un certain rapport de similitude, on ne fait que nommer le problème, non pas le résoudre ; c'est un aspect qui touche aux applications de la stratégie « mimétique » en générale, et

²⁸Ce genre d'opérateur est employé par Lewis (1978), Walton (1990), Currie (1990), et Woods (1974), pour ne prendre que quelques exemples.

²⁹L'analyse de Currie, §§2.5 et 2.6 s'appuie sur une analogie avec les inférences des croyances ; Woods s'appuie pour sa part sur la déduction plus ou moins logique (par exemple, Ch. II, §12).

³⁰C'est la démarche de Lewis, qui sera discuté prochainement.

³¹*Op cit.*, p269 *sq.*

³²Lewis (1981, p87 (§5)). Voir aussi son 1973b, §4.2.

particulièrement, comme on le montrera dans la suite, à toutes les applications à la question psychologique de la fiction.

En résumé, quelles que soient leurs autres différences, les approches de la fiction proposées jusqu'à présent souscrivent au schème général préconisé par la *stratégie* « *mimétique* ». Toutefois, la question du rapport détaillé entre fiction et réalité dans la psychologie du lecteur ne saurait pas trouver une réponse suffisante dans les théories développées en appliquant de manières diverses la stratégie « *mimétique* ». Il sera la charge de §§1.4 et 1.5 de montrer cette insuffisance. En conséquence, il faudra recourir à une *autre* stratégie. Pour ébaucher les grands traits de cette stratégie, on se sert de certains aspects du concept de jeu, lesquels, par la parenté entre le jeu et la fiction, s'appliquent également à la fiction. On introduit d'abord ces aspects, en venant par la suite à la présentation de la stratégie alternative.

1.2.2 Le jeu et la fiction

Où l'on s'appuie sur le rapprochement bien connu entre jeu et fiction pour éclaircir quelques aspects de la relation entre la fiction et la réalité à la lumière des considérations relatives au rapport entre le jeu et la réalité. Ce rapport comporte deux aspects : d'un côté, le jeu s'oppose à une partie de la réalité, dont les contraintes et les faits ne sont plus valables dans le jeu, de l'autre côté, il suppose ou se fonde sur une autre partie de la réalité, dont les faits et les contraintes sont admises, fût-il implicitement, dans le jeu. Quant à la fiction, il s'avère que les deux premières questions de §1.1.1 ont affaire à l'opposition avec la partie de la réalité qui est « annulée », alors que les deux dernières concernent le rapport avec le fond réel.

Le rapprochement entre jeu et fiction est tellement commun qu'il n'est pas besoin de le justifier, d'autant plus qu'il est invoqué par les partisans des deux traditions décrites dans le paragraphe précédent. D'un côté, Walton, comme on l'a déjà vu, conçoit la fiction en termes des jeux de *faire comme si*. De l'autre côté, Iser thématise le concept de « jeu » dans le texte³³. D'ailleurs, la structure du jeu, et notamment de son rapport à la réalité, apporte un éclaircissement particulier à la considération relative à la fiction. On reconnaît deux caractéristiques de ce rapport — d'une part, le jeu est *séparé* de la réalité — d'autre part, la liberté du joueur est contrainte non seulement par les règles internes du jeu mais

³³Selon lui, « ce n'est pas seulement que le fictif et l'imaginaire constitue le jeu : ils sont en même temps ses composants fondamentaux » (Iser, 1991, p270).

également par des *contraintes imposées par la réalité*. En un mot, le jeu s'oppose à une partie de la réalité, et il suppose ou se fonde sur une autre partie de la réalité, qui y est toujours « en vigueur », pour ainsi dire. Dans cette optique, le rapport entre la fiction et la réalité, et notamment les quatre questions posées dans §1.1.1, se présente sous un nouveau jour : les questions **A.** et **B.**, relatives aux cognitions, comportements et affects relatifs à la fiction et à leur spécificité par rapport aux cognitions, comportements et affects ordinaires, ont trait au rapport entre la fiction et la partie de la réalité qui l'oppose, alors que les questions **C.** et **D.**, relatives aux chevauchements et transferts entre fiction et réalité, ont trait au rapport entre la fiction et la partie de la réalité qui l'étaie. Cette perspective permettra, dans le paragraphe suivant, de se figurer une stratégie « non mimétique » pour aborder la fiction.

Souvent dans la littérature, on reconnaît au jeu les trois caractéristiques suivantes : un jeu est *séparé* de la vie courante – il « s'accomplit en un temps et dans un espace expressément circonscrits » ; un jeu est *réglé* – il « se déroule avec ordre selon des règles données »³⁴ ; le joueur « *est libre dans les limites des règles* »³⁵. En approfondissant un peu la relation entre le règlement du jeu et la « liberté » du joueur, on en vient à apercevoir une dimension du rapport entre le jeu et la réalité ; en considérant la relation entre le règlement et le caractère séparé du jeu, on arrive à en comprendre l'autre dimension.

Tout d'abord, pour ce qui s'agit du rapport entre le règlement et la liberté du joueur, il est question de la liberté en question relative à la manière de jouer le jeu, celle qui reste au joueur une fois qu'il a accepté de jouer selon les règles du jeu, et non celle relative à sa participation au jeu, à son acceptation même des règles. En théorie, cette liberté n'est délimitée que par les règles du jeu : hors ces règles, le joueur peut faire ce qu'il veut. Or, en pratique, elle est bornée par des contraintes supplémentaires provenant de la réalité : par exemple, les coups que les joueurs sont libres de faire ne sont pas seulement fixés par les règles du jeu (les règles du football, par exemple) mais aussi par des *contraintes réelles* sur les jeux (les faits « réels » que les joueurs ne peuvent pas survoler le terrain, lesquels ne sont pas spécifiés parmi les règles). Alors qu'il se pourrait que certaines règles du jeu entrent en conflit avec et ainsi « annulent » certaines « règles » de la réalité ou de la vie courante, il y a toujours des contraintes réelles, qui sont en l'occurrence des faits réels³⁶, qui interviennent pour limiter la liberté

³⁴Huizinga (1951, p35). Voir également Caillois (1967, p43).

³⁵Caillois, *op cit.*, p39.

³⁶Le rapport entre contrainte et fait est complexe mais bien connu ; d'ailleurs, comme l'a bien remarqué Frege, un fait (ou, dirait-il une loi) est par là même une contrainte (à penser

« absolue » laissée au joueur, en théorie, par les règles du jeu. Ces contraintes « réelles » qui s'ajoutent aux règles du jeu pour en donner le règlement « effectif » sont, dans un certains sens, *supposées* par le jeu : notamment, les règles du jeu, en précisant ce qui est interdit et autorisé dans le jeu, impliquent, quoique de manière incomplète et approximative, quelles contraintes de la vie courante sont à ignorer, et lesquelles sont toujours en vigueur.

Cette situation découle des propriétés fondamentales du rapport entre la liberté et le règlement : l'opposition entre elles – on est libre de faire ce que les règles n'interdisent pas, les règles interdisent ce qu'on n'est pas libre de faire – ne peut être pensée que dans un espace commun, et abstrait, où elle a lieu. Or, cet espace même a ses propriétés à lui ; par conséquent il exige lui-même des contraintes sur ce qu'il contient. Ces contraintes doivent être supposées lorsqu'on infère les libertés du joueur à partir des règles du jeu, elles sont des contraintes communes aux règles et aux libertés. Dans le cas des jeux, l'espace où s'opposent les libertés du joueur et les règles du jeu est un espace réel, constitué de caractéristiques de la réalité, et donc les contraintes qu'il exerce sont des contraintes réelles. Ainsi, on pourrait dire que le jeu s'érige sur un fond de contraintes réelles. Autrement dit, le jeu *suppose un fond des contraintes qu'il partage avec la réalité*.

Par ailleurs, le rapport entre le règlement du jeu et sa séparation de la réalité est un rapport d'interdépendance. D'un côté, ce sont des règles qui définissent « un espace du jeu »³⁷, un « monde temporaire au cœur du monde habituel »³⁸, par affirmation ou par implication : pour reprendre les catégories employées par Huizinga et Caillois, le lieu consacré au jeu est inscrit dans les règles qui déclarent ce qui est « hors jeu » ou non, le temps où se déroule le jeu est déterminé par les règles qui décrivent ce qui compte pour le début et pour la fin. Or, l'espace du jeu n'est pas complètement caractérisé par ses dimensions spatio-temporelles ; au contraire, les règles circonscrivent tout un espace *abstrait* du jeu, une région abstraite. Les règles des échecs, en affirmant qu'il y a un certain nombre d'entités en jeu, chacune avec certaines propriétés, décrivent un « monde abstrait », une région où se meuvent, où opèrent, où « habitent » ces entités, une région dont la limitation à un échiquier précis et un temps précis n'est qu'une projection dans le milieu spatio-temporel, voire une projection contingente. Ou encore, les jeux des enfants se déroulent dans les mondes de leurs fantaisies mais se réalisent dans un lieu et en un temps réels dont la relation avec le monde abstrait est

sa vérité). Voir par exemple Frege (1893, p xv).

³⁷Caillois, *op cit.*, p37.

³⁸Huizinga, *op cit.*, p30.

en grande partie arbitraire. Et la lecture de la fiction nous transporte dans un « univers fictionnel » qu'on reconnaît comme « un espace de jeu »³⁹ où tout se passe ; dans ce cas, « l'espace » se projette très librement dans le lieu et le temps – elle n'exige qu'un endroit quelconque et des moments quelconques pour que le jeu s'instaure.

De l'autre côté, c'est la séparation entre jeu et monde réel qui permet aux règles de s'appliquer : dans la réalité ou dans « la vie courante », ces règles sont délaissées au profit des contraintes et des faits réels. « Dans les limites du terrain du jeu, dit Huizinga⁴⁰, règne un ordre spécifique et absolu » ; hors ces limites, dans le « monde réel », les règles du jeu perdent leur sens et leur autorité. Les règles des échecs ou du football, l'identification avec des gendarmes et des voleurs, les événements de l'univers fictionnel ne tiennent pas sans les bornes abstraites qui empêchent la réalité d'entrer : bornes qui excluent le fait que ce ne soient que des figurines sur une surface, le fait que botter un ballon dans un filet n'ait pas de sens, le fait qu'il s'agisse des enfants et non pas de gendarmes ou de voleurs, le fait que ce qui est décrit ne se soit pas vraiment passé. C'est pourquoi le pire ennemi du jeu n'est pas le faux joueur mais le « briseur du jeu » car c'est lui qui franchit les frontières abstraites du jeu est faire entrer la réalité ; c'est lui, comme dit Huizinga⁴¹, qui « brise son monde magique ». Le jeu *s'oppose donc à la réalité*, dans la mesure où lui est nécessairement séparé.

Afin de pouvoir mettre en rapport aussi bien l'opposition du jeu par rapport à la réalité (des faits et des contraintes réelles sont « annulés » par les règles du jeu) que son appui sur elle (des contraintes réelles toujours effectives dans le jeu), il faut distinguer deux sens d'espace. Pour ce qui concerne son opposition à la réalité, « l'espace du jeu » est un *milieu* désigné, une *région* circonscrite par des règles, une *place* distincte de celle de la réalité ; en revanche, les contraintes réelles qui s'imposent dans le jeu sont conçues comme des caractéristiques d'un espace sous-jacent au jeu, qui le sert de *fond*, où espace est entendu ici au sens *d'étendue* indéfinie. Or ce fond réel est commun au jeu et à la partie de la réalité qui s'y oppose ; autrement dit, la partie de la réalité à laquelle s'oppose le jeu est une région dans l'espace qui sert de fond pour le jeu. En conséquence, l'opposition entre jeu et réalité se pense naturellement comme ayant lieu dans l'espace constitué par les contraintes réelles qui sont impliquées dans le jeu. Dans cette optique, les règles du jeu délimitent une région dans un espace abstrait ; le réel est la structure même de cet espace, mais elle est également une région

³⁹Par exemple, Schaeffer (1999, p176).

⁴⁰Huizinga, *op cit.*, p30.

⁴¹Huizinga, *op cit.*, p32.

dans cet espace qui est distincte de la région du jeu. Les contraintes et les faits réels qui sont *supposés* par le jeu sont celles qui *déterminent l'espace abstrait dans lequel le jeu se définit* ; les faits et les contraintes réels qui sont *annulés* dans le jeu sont celles qui *définissent la région du réel dans cet espace, région à laquelle la région du jeu s'oppose*. Dans les échecs, les coups possibles ont un sens seulement dans un espace abstrait où sont respectées certaines lois logiques et topologiques – le fond commun à la réalité et au jeu – ; or dans cet espace les possibilités des coups définissent une région restreinte qui se distingue de la région réelle par la pauvreté et la rigidité des coups possibles – dans la réalité ces règles ne s'appliquent pas et c'est ainsi que le jeu se sépare de la réalité. Puisque ce sont les règles du jeu qui impliquent, de manière approximative et incomplète, voire même instable et variable selon l'interprétation du jeu, quelle partie de la réalité est à supposer dans le jeu et laquelle est à opposer au jeu, ce sont des règles du jeu qui exigent cette topologie. On pourrait donc dire que les règles esquissent une sorte de « nivellation » locale autour d'elles : les contraintes de la réalité qui se trouvent sur le même niveau que les règles du jeu entrent en opposition avec le jeu et sont annulées dans le jeu , alors que celles qui se trouvent sur un niveau plus basique sont maintenues voire supposées dans le jeu.

En conjuguant cette perspective sur les jeux avec le rapprochement entre fiction et jeu, on arrive à un nouvel aperçu sur les structures relatives à la fiction qui sont identifiées par les questions de §1.1 et accentuées par divers philosophes. L'univers fictionnel qui se distingue du monde réel se rapproche d'un « espace de jeu » ; on comprend maintenant que ce sont les « règles de la fiction » qui à la fois soutiennent cet espace et dépendent de lui. Comme l'a anticipé Caillois et l'a affirmé Walton, la règle de la fiction est celle du *comme si*⁴² : pour « jouer le jeu », on doit « imaginer » les énoncés du texte ou « *faire comme si* » ils étaient vrais, et réciproquement, le fait d'imaginer ainsi implique la création d'une région abstraite (« espace de jeu », « univers fictionnel »), qui s'oppose au « monde réel », où se déroule le *faire comme si*. La question **A.** de §1.1.1 a trait à ces aspects.

En outre, puisque l'opposition entre l'univers fictionnel et le monde réel n'a lieu que dans un espace déjà réel, ce qui a été appelé le fond réel du jeu, les « transferts » entre fiction et réalité auxquels ont trait les questions **C.** et **D.** sont reconnus avoir affaire au fond commun à la fiction et au réel.

⁴²Caillois, *op cit.*, p40 : « le sentiment de *comme si* remplace la règle et remplit exactement la même fonction » ; Walton (1990, p40) : « Le fait pour un énoncé d'être fictionnel (*fictionality*) [...] consiste dans le fait qu'une règle du jeu prescrit que l'on l'imagine ».

D'un côté, les entrées de la réalité dans la fiction repérées par la question **C.** trouvent un équivalent dans le jeu : dans le jeu d'échecs, il faut s'appuyer sur certaines propriétés topologiques et logiques de base ; dans les jeux d'enfants, la connaissance des autres joueurs et de leurs propriétés (qui est « rapide » ou non) réelles peuvent intervenir. Ces aspects du réel à qui il est permis d'entrer dans le jeu sont les éléments du « fond réel » commun au jeu et aux aspects de la réalité auxquels il s'oppose.

De l'autre côté, les effets et les influences de la fiction sur la réalité auxquels réfère la question **D.**, trouvent eux aussi un analogue dans les jeux. A proprement parler, les notions d'échec et mat, de marquer un but, d'arrêter un autre enfant n'ont pas de sens hors les jeux respectifs ; néanmoins on en parle en dehors des jeux, de même qu'on parle des personnages fictionnels hors de l'immersion fictionnelle. La thèse de Huizinga – que les notions telles que la vertu et l'honneur se définissent, s'attribuent et proviennent en grande partie des jeux⁴³ – constitue un exemple plus intéressant. Ces notions-ci ne sont pas seulement attribuées et considérées dans le jeu ; au contraire, leur importance dans la réalité constitue un exemple des « transferts » du jeu dans la réalité⁴⁴. On entend retentir ici les thèses de Goodman⁴⁵ relatives à l'applicabilité réelle de prédicats saisis d'abord dans la fiction, ou de Ricoeur⁴⁶, pour ne prendre qu'un seul exemple, sur la teneur morale de la fiction, et de l'importance de cette caractéristique. Il s'agit encore des rapports entre le jeu et son fond réel : dans le jeu (et donc dans la fiction) peuvent apparaître ou être « soulignées » des propriétés de l'espace abstrait sous-jacent au jeu (par exemple, la propriété de pouvoir accueillir les éléments du jeu, telles les notions de l'échec et mat, du but, de l'arrestation, des personnages fictionnels, munis de certaines propriétés, telles que la vertu et l'honneur). Or, puisque cet espace est commun au jeu et au réel, puisqu'il est un espace réel, *une dimension du réel*, les propriétés qui sont « découvertes » dans le jeu sont des « propriétés réelles », de sorte qu'elles demeurent même après la disparition du jeu. Le transfert du jeu à la réalité se comprend comme le repérage dans le jeu des faits relatifs à l'espace sous-jacent réel : inversement, la question de ces transferts est reliée à la question du fond commun, et aux modalités de sa détermination (incomplète) par les règles du

⁴³Huizinga, *op cit.*, par exemple p112 : « Vertu, honneur, noblesse et gloire se trouvent ainsi dès le début dans le cadre de la compétition, c'est-à-dire du jeu. »

⁴⁴Il ne s'agit pas ici de souscrire à la thèse d'Huizinga, mais seulement de l'employer comme exemple.

⁴⁵Goodman, *op cit.*, §VI.5.

⁴⁶Ricoeur, *op cit.*, p113-117. Ce rapprochement entre la fiction et le moral, apparenté au rapprochement entre l'art et le moral, connaît une longue histoire, remontant au moins jusqu'à Hegel, voire Platon.

jeu.

La conception du jeu développée ci-dessus permet donc de comprendre l'entière du rapport entre fiction et réalité – l'opposition entre fiction et réalité sur fond d'un espace qui est lui-même réel – aussi bien que les questions posées à son égard. Pour dire vite, les questions **A.** et **B.** ont affaire à l'opposition, alors que les questions **C.** et **D.** ont affaire au rapport entre fiction et son fond réel. Cette structure est entièrement générale, s'appliquant de fait dans un grand nombre de cas ou de « domaines » différents, ayant trait aussi bien aux conceptions sociales impliquées dans la fiction qu'à la notion de vérité que lui est associée, aussi bien aux « entités » de la fiction qu'aux activités, du lecteur ainsi que de l'auteur, relatives à la fiction. Pour le propos de ce travail, il importe seulement que, comme on le montrera dans la suite, cette structure s'applique à la psychologie du lecteur.

Non seulement ce détour par les jeux aboutit-il à une perspective différente du problème entier, à savoir celle du rapport entre la fiction et la réalité, mais de plus, il permet de bien apercevoir une *autre* stratégie pour aborder la question générale de ce rapport (dans la psychologie du lecteur). On se charge, dans le paragraphe suivant, d'explicitier maintenant l'idée de cette stratégie.

1.2.3 La stratégie adoptée

Où l'on esquisse, à grands traits, la stratégie qui sera adoptée pour développer, dans les Chapitres 2 et 3, une théorie de la lecture de la fiction. Il s'agit de partir d'un niveau de base où la distinction entre fiction et réalité n'est pas donnée ni supposée, pour ensuite retrouver cette distinction et le rapport entre les deux termes.

L'image à laquelle on est arrivé est la suivante. La fiction occupe une aire dans un espace général que l'on qualifie de « réel » (ayant trait à ou provenant de la « réalité ») ; dans cet espace, elle s'oppose à une autre aire qui pourrait également qualifiée de « réel ». La stratégie « mimétique » aborde cette situation de manière suivante : elle repère d'abord l'aire « fictionnelle » et *une* partie de la réalité – tantôt l'aire qui s'oppose à la fiction tantôt l'espace qui l'étaie (tantôt des aspects des deux) – pour ensuite considérer les rapports entre elles. La stratégie dont on se servira dans les chapitres suivants est légèrement différente. Elle *ne commence pas* par *poser* ou par *repérer* deux choses – fiction et réalité – pour *ensuite* les mettre en rapport ; elle *part* plutôt d'un *espace général de base* pour *retrouver ensuite* l'aire fictionnelle, l'aire réelle et *en même temps* le rapport entre ces deux aires et l'espace de base. Dans la mesure où, au niveau

de l'espace de base, du fond réel, il n'y a pas d'opposition entre fiction et réalité, la stratégie ne *pose* pas la distinction entre fiction et réalité, elle la *retrouve*.

Deux éventuels écueils peuvent se présenter à une telle stratégie : d'une part, vu que les règles du « jeu habituel » ne déterminent pas la distinction entre la partie de la réalité qui s'oppose à la fiction et la partie qui l'étaie *de manière complète*, voire même *stable*, il pourrait être difficile de trouver un « espace réel » stable qui est toujours approprié à étayer la fiction et la partie de la réalité à laquelle elle s'oppose. D'autre part, même si l'on peut trouver un tel espace réel, il n'est pas évident que l'on puisse retrouver l'aire fictionnelle, l'aire réelle et le rapport entre elles à partir de lui. Il n'est certainement pas toujours possible de contourner ces écueils. Il appartient à l'analyse proposée de montrer comment il est néanmoins possible, voire désirable, de le faire dans le cas spécifique de la psychologie du lecteur.

Quant au premier doute, à l'égard de l'existence d'une base suffisante, on argue que, au lieu de partir de la distinction entre deux genres d'états psychologiques – la croyance (usuelle) et le *faire comme si* – comme le préconise la stratégie « mimétique » dans l'application proposée par Walton et Currie, il est possible de spécifier un niveau psychologique plus général et fondamental, où les propriétés et les aspects importants sont tellement généraux à être communs à la fiction et à la réalité. En l'occurrence, ce niveau comprend les états de la conscience : bien évidemment, les réflexions du sujet, ses attitudes et ses comportements, qu'ils soient relatifs à la fiction ou à la réalité, passent *d'une manière ou d'une autre* par la conscience.

Par ailleurs, on répond au second doute, ayant trait à la difficulté de retrouver la fiction et la réalité, en mobilisant l'hypothèse de travail explicitée dans §1.1.2, selon laquelle le lecteur sait que le texte qu'il lit est fictionnel. Cette hypothèse n'implique aucune attitude mentale ni état psychologique spécifique à la fiction ; au contraire, elle ne fait que poser une connaissance particulière comme aspect de la psychologie du lecteur, à même titre que toute autre connaissance. On retrouve les aspects psychologiques relatifs à la fiction comme les comportements, les conduites, les configurations d'états psychologiques que l'on associe à cette connaissance, dont toutes sont perceptibles à un niveau psychologique « de base » où la distinction entre fiction et réalité ne figure pas *de prime abord*. Comme on l'a dit dans §1.1.2, on retrouve la fiction et la réalité, aussi bien que le rapport entre elles, en tant qu'elles apparaissent dans la psychologie du lecteur, comme des *conséquences* de la connaissance qu'un texte est fictionnel. On pourrait peut-être qualifier cette approche de « *naturaliste* », dans la mesure où elle se sert seulement des états psychologiques bien connus et d'une connaissance

universellement admise pour aborder la fiction et la question de son rapport à la réalité dans la psychologie du lecteur, sans à aucun moment *poser* ou *supposer* des aspects, entités ou notions spécifiques à elle.

La démarche est la suivante. Après avoir, dans §1.3, fait quelques considérations préalables sur la catégorisation des théories du mental en général, on mettra en question, dans §§1.4 et 1.5, les approches de la question « mentale » de la fiction qui emploient toutes la stratégie « mimétique ». Au cours de cette critique, les questions épineuses de la fiction seront reconnues comme relatives à un aspect psychologique « de base » (grossièrement, la conscience). En prenant la voie suggérée par cette critique, il s'agira dans les chapitres suivants (Chs. 2 et 3) d'aborder la psychologie du lecteur – ses aspects « réels » aussi bien que « fictionnels » – à partir de ce niveau « de base ». Ce qui permettra une réponse, ou du moins ouvrira une voie à la réponse, aux questions détaillées de §1.1.

1.3 Approches psychologiques à l'égard de la fiction

Les divers auteurs qui ont songé à la question psychologique de la fiction posée dans §1.1 – la question du rapport entre fiction et réalité dans la psychologie du lecteur – ont, pour l'essentiel, abordé cette question selon la stratégie « mimétique ». C'est-à-dire qu'ils posent d'abord une distinction entre « quelque chose de mental » qui est relatif à la fiction et « quelque chose de mental » qui est propre à la réalité (ou la vie ordinaire), pour ensuite poser ou soutenir certaines ressemblances entre les deux « choses » distinguées. Ce genre de stratégie a été employée non seulement avec les concepts utilisés pour parler des la psychologie des êtres humains (le *faire comme si* de Currie, l'imagination de Walton), mais aussi avec les mécanismes mentaux dont on soutient qu'ils sont « derrière » leurs comportements. Il s'agira, dans §§1.4 et 1.5, de considérer en détail certaines théories qui appliquent cette stratégie dans divers domaines de la psychologie du lecteur et de diverses manières, pour mettre en évidence certaines faiblesses communes à toutes. Ce faisant, on motivera une autre stratégie. Pour bien préparer cette discussion, il est nécessaire de rappeler quelques distinctions bien connues dans la philosophie de l'esprit qui peuvent, par leur absence, donner lieu à des confusions nuisibles. On va illustrer ces aspects sur des théories existantes de la psychologie de la fiction.

On rappelle premièrement la distinction entre d'une part des « aspects conceptuels », ayant trait aux notions psychologiques qu'on emploie dans les attribu-

tions des états psychologiques aux sujets, et d'autre part des « aspects causaux », ayant trait aux mécanismes, normalement internes, qui sous-tendent les comportements manifestes (§1.3.1). Deuxièmement, il faut reconnaître la distinction entre les notions psychologiques ordinaires, employées quotidiennement, et les notions psychologiques plus sophistiquées, qui représentent des raffinements plus ou moins féconds des premières (§1.3.2). Enfin, on fait quelques remarques sur les problèmes particuliers que posent les états affectifs (§1.3.4) ; ces remarques s'appuient sur une distinction introduite préalablement (§1.3.3) entre les États Mentaux qui peuvent être attribués sur la base de l'état instantané du sujet, tel peut-être le sentiment, et ceux dont l'attribution dépend nécessairement de plusieurs moments, dont supposément la croyance et l'émotion.

1.3.1 La question « conceptuelle » et la question « causale »

Où l'on rappelle la distinction, bien connue dans la philosophie de l'esprit, entre les aspects dits « causaux », tels les processus mentaux qui sous-tendent les comportements du sujet, et les aspects dits « conceptuels », c'est-à-dire les concepts employés pour décrire et comprendre ces comportements.

Apparemment, on pensait jadis que toutes les attributions d'états psychologiques à un sujet correspondent à quelque chose d'effectif « dans sa tête ». Or, au cours du dernier siècle notamment, on a appris à distinguer entre, d'une part ce qu'on dit (ordinairement) au sujet de la psychologie d'un agent et, d'autre part, les processus mentaux qui « sous-tendent » ou « produisent » effectivement les comportements manifestes de l'agent. C'est effectivement la distinction que tire Fodor entre les aspects « conceptuels » et les aspects « causaux » d'une question⁴⁷ : les aspects « causaux » sont les mécanismes et le système qui donnent lieu à certains comportements ; les aspects « conceptuels » sont les notions et les concepts mis en œuvre pour comprendre et décrire ces comportements. Ou, autrement, la distinction de Dennett entre la perspective interprétative (*intentional stance*), qui a affaire aux attributions psychologiques à l'agent, et la perspective planificatrice (*design stance*)⁴⁸, qui a trait plutôt aux mécanismes et aux structures derrière les comportements. Malgré les diverses positions sur le rapport entre les deux aspects ou « niveaux » distinguées, on s'accorde actuellement sur le fait que la distinction est fondée.

⁴⁷Fodor (1975, p6-7).

⁴⁸Dennett (1987, Ch 2 et 3). Ce ne sont pas les traductions employées dans la traduction française.

On pourrait illustrer cette distinction sur quelques théories proposées relativement à la fiction. D'une part, l'imagination de Walton et le *faire comme si* de Currie (§1.2.1) sont attribués au lecteur pour interpréter ou comprendre ses comportements : ce sont des *concepts* suggérés en réponse à la question psychologique « conceptuelle » relative à la fiction. Il y a cependant une différence importante entre eux : le *faire comme si* de Currie est une *attitude propositionnelle*, dans la mesure où elle prend comme argument une proposition (« Sherlock Holmes est un détective »), alors que l'imagination de Walton ne l'est pas nécessairement, dans la mesure où elle pourrait prendre d'autres types d'arguments (on peut imaginer être Napoléon)⁴⁹. Pour désigner les « concepts mentaux » qui prennent comme argument des propositions aussi bien que ceux qui peuvent ne pas les prendre, il faut éviter le terme « attitude », car il est souvent employé dans la philosophie analytique comme ellipse pour « attitude propositionnelle » ; on adoptera plutôt l'expression « *État Mental* ». Parmi les États Mentaux, on compte les croyances, les désirs, les *faire comme si*, les imaginations, mais également les émotions : bref, les concepts que l'on emploie pour parler des aspects plus ou moins mentaux d'êtres humains. L'imagination de Walton et le *faire comme si* de Currie sont des États Mentaux employés pour comprendre et décrire le comportement du sujet : on comprend le fait qu'il dit que Sherlock Holmes est un détective, sans penser à l'appeler en cas de crime, par l'attribution de l'État Mental approprié – il *imagine que* (ou *fait comme si*) Sherlock Holmes est un détective, ce qui a des conséquences sur son comportement.

D'autre part, Currie approfondit sa théorie de l'État Mental « *faire comme si* » en proposant qu'il y a une sorte de *mécanisme*, qu'il appelle « simulateur », dans la tête du lecteur qui simule certains fonctionnements ordinaires du mental, sur les opérandes et les résultats qui ne sont pas des croyances réelles mais des « *faire comme si* »⁵⁰. Pour lui, l'attitude d'imaginer ou de « *faire comme si* » que l'on attribue au lecteur correspond à ce « mécanisme » : « l'imagination est le simulateur »⁵¹. Cette thèse n'a pas affaire à l'attribution d'un concept de *faire comme si* au sujet, mais plutôt à une « cause » ou un processus effectif qui sous-tend supposément les comportements du lecteur. Il s'agit d'un aspect « causal ».

L'intérêt des théories proposant des mécanismes mentaux relatifs à la fiction

⁴⁹Currie, *op cit.*, §2.4 ; Walton, *op cit.*, §1.4 et pp42-43. Il n'est pas nécessaire de s'attarder ici sur la question de savoir si les attitudes *de se* (au sujet de soi-même) sont des attitudes propositionnelles.

⁵⁰Currie (1995). Voir §1.5.1 ci-dessous.

⁵¹Currie, *op cit.*, p 158.

ne tient cependant pas à l'*effectivité* des mécanismes proposés, mais plutôt à leur *efficacité* pour rendre compte des compétences et des comportements du lecteur relatifs à la fiction. En effet, le fait d'adopter le cadre « internaliste » a pour conséquence que les questions de l'effectivité ou du réel sont *mises entre parenthèses* (§1.1.2), à moins qu'elles soient comprises comme questions relatives au réel en tant que le sujet se le représente. Il n'est certainement pas question ici, comme il l'est pour les psychologues cognitivistes, des *vrais mécanismes* qui sont *réalisés physiquement* par le cerveau humain. En revanche, l'importance des mécanismes qui sont proposés par les psychologues cognitivistes pour le propos de ce travail tient à leur *structure* plus *fine*. Alors que les États Mentaux sont des concepts employés ordinairement pour comprendre *certaines* comportements de l'agent (les plus « importants » ou « intéressants »), les mécanismes et les processus « causaux » proposés par les psychologues cognitivistes devraient rendre compte de *tout* son comportement ; par conséquence, ils devraient offrir une théorie plus riche de son mental. D'ailleurs, on verra dans §1.4 que les États Mentaux ne suffisent pas pour aborder les questions détaillées de la fiction, d'où un recours, dans §1.5, aux mécanismes psychologiques. Là, on s'intéressera, pour répéter le point, aux *structures* de ces mécanismes, et à la possibilité d'aborder des questions de la fiction en termes de ces structures, non pas à leur « réalisme ».

La distinction entre États Mentaux et mécanismes sert à organiser la discussion suivante : dans un premier temps (§1.4), on discute les théories proposant des *États Mentaux* appropriés à la fiction, pour venir, dans un deuxième temps (§1.5), aux *mécanismes* « causaux » qui ont été proposés.

Or, quant aux théories des aspects « conceptuels » du mental, c'est-à-dire des théories d'États Mentaux, on pourrait distinguer approximativement deux *conceptions* d'États Mentaux qui sont proposés dans la littérature. La discussion des États Mentaux spécifiques à la fiction (« imagination », *faire comme si*) s'ouvrira à la confusion si l'on n'est pas clair dès le départ sur la conception d'État Mental en jeu. Il appartient au prochain paragraphe de présenter cette distinction.

1.3.2 Conceptions ordinaires et sophistiquées des États Mentaux

Où l'on expose la distinction entre deux conceptions des États Mentaux. Comme préliminaire, on introduit la notion de bloc d'attribution, comportant tous les États Mentaux qui sont attribuées d'emblée,

avec les expériences et les comportements auxquels ils réfèrent et les rapports appropriés entre eux. La conception ordinaire, telle celle que l'on emploie quotidiennement, est caractérisée d'une part, par le fait que le bloc d'attribution de l'État Mental ne doit pas comporter des États Mentaux relatifs à tous les énoncés du langage – il est non saturé – et, d'autre part, par le fait que son bloc d'attribution contient des comportements et des expériences ayant lieu dans une étendue temporelle restreinte – sa base d'attribution est locale. Par contraste, la conception sophistiquée, telle celle qui est à l'œuvre dans la théorie de la décision, propose des États Mentaux qui sont attribués dans des blocs où il y a un État Mental à l'égard de chaque énoncé du langage – ils sont saturés – et où figure tout comportement et expérience du sujet, réel ou contrefactuel – sa base d'attribution est globale. Il s'avère que, quelles que soient leur prétentions, le faire comme si en tant qu'il est proposé par Walton et Currie, est un État Mental ordinaire.

Les États Mentaux d'imagination ou de *faire comme si* sont censément analogues aux États Mentaux usuels, notamment à la croyance (§1.2.1). Or, il y a, sur le marché si l'on peut dire, deux *conceptions* différentes des États Mentaux usuels, et notamment de la croyance, de telle sorte que la question se pose de savoir à *quelle* conception les États Mentaux d'imagination ou de *faire comme si* sont analogues. Alors que cette question n'est pas discutée par Walton et Currie, elle est importante pour la discussion critique de leur théorie. En fait, la critique tient quelle que soit la conception d'État Mental employée, mais elle prend, pour différentes conceptions, des formes légèrement différentes. On expose ici deux aspects qui sont communs aux deux conceptions d'États Mentaux et deux points qui les distinguent, pour pouvoir ensuite caractériser le *faire comme si* de Currie et Walton.

Une propriété qui est commune à toutes les approches du mental dans la tradition analytique récente, qu'il s'agisse d'une des deux conceptions des États Mentaux que l'on mettra en évidence ou des théories de la psychologie cognitive qui posent des mécanismes psychologiques, concerne les « faits » qu'elles reconnaissent : les seules « *data* » ou données que ces théories se permettent pour l'attribution des États Mentaux à l'agent (ou dans l'explicitation d'une théorie psychologique) consistent en des observations à la troisième personne de son comportement dans des circonstances particulières. Les expériences phénoménologiques, saisies à la première personne, n'en font pas partie. On appelle cette contrainte méthodologique sur l'ensemble de données que les théories

se donnent « l'*externalisme méthodologique* », à défaut d'un meilleur terme⁵². Par exemple, on attribue à quelqu'un la croyance qu'un parapluie protège de la pluie sur la base de son comportement (lorsqu'il sort dans des conditions pluvieuses, il prend un parapluie) plutôt que sur la base ses expériences phénoménales.

Un deuxième point commun aux théories des États Mentaux tient aux modalités d'attribution de ces États : une attribution des États Mentaux se fait de manière générale dans un « bloc » plus ou moins grand, comportant éventuellement plusieurs États Mentaux et faisant référence à plusieurs comportements et plusieurs expériences de l'agent. Souvent, on n'attribue pas isolément la croyance relative aux propriétés protectrices des parapluies, on y joint par exemple l'attribution du désir de ne pas se mouiller. On parlera dans la suite de *blocs d'attribution*, lesquels comportent des États Mentaux attribués ensembles, des expériences et des comportements appropriés de l'agent, reliés par les rapports appropriés⁵³. Un bloc consiste par exemple en une croyance à des propriétés protectrices du parapluie, un désir de ne pas se mouiller, l'action de prendre un parapluie lorsqu'il pleut, avec les rapports « d'explication » de cette action en termes de ces États Mentaux. Ce que l'on appelle ici bloc d'attribution est donc apparenté à la « théorie » comportant, comme termes théoriques, des États Mentaux, laquelle, selon Lewis, est en jeu dans l'attribution des États Mentaux à un sujet⁵⁴. Pareillement aux termes théoriques, un État Mental attribué dans

⁵²Peut-être la notion la plus proche dans la littérature est-elle le « behaviorisme méthodologique » au sens de Sellars (1997, par exemple p100) : « L'exigence behavioriste [est] que tous les concepts doivent être introduits dans des termes relatifs au comportement manifeste ». Pourtant la terminologie dans le domaine est devenue tellement surchargée qu'il n'est plus sûr que le sens sellarsien de « behaviorisme méthodologique » soit courant : pour prendre un seul exemple, Graham (2005) identifie le behaviorisme méthodologique comme « l'affirmation selon laquelle la psychologie devrait s'occuper du comportement des organismes... et non pas des états ou événements mentaux » (il attribue cette position à Watson, par exemple 1913). En outre, on entend par « externalisme méthodologique » ici la doctrine méthodologique qui permet toute donnée qui est observée à la troisième personne, y compris les données neurophysiologiques telle l'imagerie cérébrale. Cela n'importe pas pour les théories des États Mentaux, mais elle permettra d'appliquer ce terme aux théories proposant des mécanismes psychologiques.

⁵³La terminologie « bloc d'attribution » vise également à être neutre par rapport au statut ontologique et métaphysique des États Mentaux ; comme l'est d'ailleurs exigé par la perspective « internaliste » de ce travail (§1.1.2), on ne prend aucun côté (réaliste, instrumentaliste, anti-réaliste, etc.) dans ces genres de débat.

⁵⁴Lewis (1972, notamment §III) nous invite à « penser la psychologie ordinaire comme une théorie scientifique qui introduit des termes » (voir aussi sa note 12). Soulignons qu'il ne s'agit pas de se limiter au fonctionnalisme de Lewis, mais d'employer la notion de bloc – ou de théorie – comme schéma général couvrant toutes les positions contemporaines à l'égard des États Mentaux. Le behaviorisme logique, par exemple, attribue des États Mentaux en blocs,

un bloc est relatif à ce bloc, de telle sorte que non seulement les propriétés de l'État Mental attribué sont héritées de sa position dans le bloc (relative à d'autres États Mentaux et comportements en jeu dans l'attribution), mais de plus on ne peut pas l'employer ou en parler hors de ce bloc, quitte à l'accorder un sens ou un usage différent. Par ailleurs, les blocs d'attribution peuvent différer selon la taille, comportant plus ou moins d'États Mentaux – éventuellement un seul – et plus ou moins de stimulations et de comportements de l'agent. On arrive à caractériser la distinction entre les deux conceptions d'États Mentaux dont il est question en termes des propriétés des *blocs* auxquels ils appartiennent et par ou dans lesquels ils sont attribués.

De manière intuitive, il y a apparemment une différence entre deux conceptions d'États Mentaux qui sont discutés la littérature : l'une, que l'on peut qualifier d'*ordinaire*, est celle que l'on emploie dans la vie quotidienne ; l'autre, que l'on peut appeler *sophistiquée*, est développée par des théories plus ou moins rigoureuses, dont notamment la théorie de la décision. On parlera des *États Mentaux ordinaires* et des *États Mentaux sophistiqués*, plutôt que des conceptions ordinaires (respectivement, sophistiquées) des États Mentaux. L'exemple de la croyance aux propriétés protectrices des parapluies permet d'illustrer la différence entre les deux conceptions. Aussi bien la conception ordinaire employée dans la vie quotidienne et étudiée par les philosophes intéressés à l'usage quotidien du langage, dont notamment Ryle⁵⁵, que la conception sophistiquée, développée à l'instar de Ramsey, rapportent la croyance à quelque chose de l'ordre de la tendance, de la disposition ou de la propension⁵⁶, en l'occurrence la tendance à prendre le parapluie lorsqu'il pleut. Or, alors que la conception ordinaire garde un sens *non rigoureux* et *vague* de cette tendance et donc de la croyance qui lui est associée⁵⁷, la conception sophistiquée en donne un sens

mais avec une structure différente (chaque État Mental a une définition explicite en termes des comportements du bloc).

⁵⁵ « Il n'y a pas de manuels de la psychologie ordinaire (bien que le *Concept d'Esprit* de Ryle pourrait être employé en dernier recours) » Dennett (1987, p46).

⁵⁶ Ryle (2000, p114) : « Les termes dispositionnels tels 'connaître', 'croire', ... signifient des capacités, des tendances ou des inclinations à faire, non pas des choses d'un seul et unique genre, mais des choses de plusieurs genres » ; Ramsey (2003, p164) : « le degré de croyance est une propriété causale de la croyance, qu'on peut vaguement définir comme la propension d'agir sur la base de cette croyance ».

⁵⁷ « 'Savoir le français' est une phrase vague mais son imprécision ne la rend pas inutile pour autant », Ryle, *op cit.*, p119. On remarque que la complexité et le caractère vague de la notion de tendance employée (voir également la citation dans la note 56) pourraient permettre aux utilisateurs de cette conception de penser une croyance comme tendance à faire une action sans expliciter le désir de faire cette action (dans l'exemple du parapluie, le désir de ne pas mouiller). De même qu'en physique on peut dire que l'accélération d'un corps

rigoureux et technique en tant que nombre réel entre zéro et un – le dit « degré » de la croyance – qui entre dans les « calculs » précis pour donner les comportements de l'agent dans des circonstances différentes⁵⁸.

Hormis les évidentes différences intuitives entre les deux conceptions, on pourrait mettre en avant deux aspects qui permettent une caractérisation plus précise de la distinction entre elles et qui seront importantes dans la suite, alors qu'ils ne sont pas toujours explicités dans la littérature. A savoir, les États Mentaux sophistiqués sont *saturés*, alors que les États Mentaux ordinaires ne le sont pas ; et les États Mentaux sophistiqués ont des *bases d'attribution globales*, alors que les États Mentaux ordinaires ont des *bases d'attributions locales*.

En premier lieu, un bloc d'attribution est dit *saturé* relativement à un langage donné⁵⁹ s'il comporte des États Mentaux relatifs à chaque énoncé du langage, et *non saturé* si ce n'est pas le cas. Donc, dans la mesure où la théorie proposée par Ramsey accorde à tout énoncé un degré de croyance (aussi bien, par ailleurs, qu'une valeur de préférence)⁶⁰, elle propose des blocs *saturés* pour attribuer les États Mentaux : on dira que la conception d'États Mentaux proposée est *saturée*. En revanche, dans les attributions des États Mentaux à un individu dans la vie quotidienne, d'autres États Mentaux ne sont pas nécessairement en jeu : lorsque l'on attribue à quelqu'un, en voyant sa tendance à prendre un parapluie dans la pluie, la croyance que le parapluie protège de la pluie, il n'y a aucune question de déterminer son état de croyance relative au fait que Jacques Chirac est le président de la France. Le bloc d'attribution accorde un statut de croyance à un énoncé (« il pleut ») mais non pas à l'autre (« Jacques Chirac est président de la France ») et donc n'est donc nullement saturé. On dira que les États Mentaux ordinaires sont *non saturés*.

En second lieu, on appelle *base d'une attribution* (ou d'un bloc d'attribution) la collection d'expériences et de comportements qui figurent dans le bloc d'attribution approprié, c'est-à-dire les « données » selon l'externalisme méthodologique. Une attribution d'États Mentaux se fait sur une base *locale* si les expériences et les comportements de cette base n'occupent qu'une *étendue*

est proportionnelle à la force, *en supposant la masse fixe*, on pourrait identifier une croyance à une tendance à une certaine action, *en supposant certains désirs et autres croyances fixes*. Puisqu'une tendance n'est pas une nécessité stricte, on pourrait rapprocher les cas où ces désirs ou autres croyances manquent aux exceptions de la tendance, d'où le vague et la complexité de la notion de tendance, requis pour pouvoir admettre les multiples « sortes » d'exceptions.

⁵⁸Ramsey, « Vérité et probabilité », dans 2003, notamment §3.

⁵⁹On suppose un langage fixe pour le reste de cette partie, et donc ne l'explicite pas dans la discussion ultérieure.

⁶⁰*Op cit.*, p167. Les énoncés que le sujet admet comme certainement faux reçoivent une utilité zéro : en termes ramseyiens, les mondes où ces énoncés sont vrais ont une utilité zéro.

temporelle restreinte ; d'autre part, l'attribution se fait sur une base *globale* si cette base de comportements et d'expériences du sujet est « maximale », comportant *tous* les comportements du sujet pour *la totalité de l'étendue de sa vie*, ainsi éventuellement que *ses comportements* « *contrefactuels* », c'est-à-dire des comportements qui interviendraient si le sujet se trouvait dans des circonstances différentes. Dans la vie ordinaire, on n'a affaire qu'à des attributions de base locale : si l'on attribue à quelqu'un la croyance à l'imminence de la pluie, on ne le fait que sur la base de quelques stimulations (il regarde sur la fenêtre) et de quelques comportements (il prend son parapluie pour sortir) de peu d'étendue, et dont d'autres stimulations (il regarde son repas) et comportements (il le mange) sont absents. En revanche, la théorie de la décision tend à prendre une base d'attribution globale, le problème « d'Interprétation Radicale » de Lewis, qui consiste *grosso modo* à donner l'ensemble de toutes les croyances et les désirs d'un agent, étant donné tous ses comportements (et expériences) effectifs et contrefactuels⁶¹, étant à cet égard l'exemple le plus remarquable. La taille impressionnante de la base n'implique cependant pas que les États Mentaux attribués par ce genre de théorie soient constants, car on étiquette les croyances et les désirs par le temps approprié et on envisage des règles pour régir les changements des croyances entre des temps différents, telle la conditionalisation bayésienne.

Vu ces différentes conceptions d'États Mentaux, il est légitime de se demander à quelle conception souscrit l'État Mental de *faire comme si*, en tant qu'il est proposé par Walton et Currie. Le *faire comme si* qu'ils emploient, est-il un État Mental ordinaire, c'est-à-dire non saturé, avec base d'application locale, ou sophistiqué, c'est-à-dire saturé, avec base d'application globale ?

Il semble à première vue que les deux auteurs, et notamment Currie, aient en tête la conception *sophistiquée* des États Mentaux : dans la mesure où la structure des « croyances dans le jeu de *faire comme si* » (« make-beliefs ») et des « désirs dans le jeu de *faire comme si* » (« make-desires ») que propose Currie

⁶¹Lewis (1974, p 109) : sa base de données « nous dit non seulement l'état physique actuel [du sujet] mais son histoire physique, non seulement les faits physiques particuliers actuels mais également les dépendances nomiques (*nomic*), causales et contrefactuelles entre eux ». Ces dépendances contrefactuelles doivent inclure les comportements que le sujet *effectuerait* si ses conditions physiques n'étaient pas ce qu'ils sont effectivement, c'est-à-dire ses comportements contrefactuels. Même s'il n'est pas thématiqué, l'appui sur une base globale est déjà présent chez Ramsey, où l'attribution d'une croyance au sujet s'appuie sur sa réponse à certains paris (*op cit.*, p169 *sq.*), de sorte que pour attribuer des croyances relatives à tous les énoncés du langage, comme l'exige le caractère saturé de la théorie de Ramsey, il faut offrir de tels paris pour chaque énoncé du langage : puisque cela n'arrive pas effectivement dans la vie du sujet, on a plutôt trait aux offres contrefactuels au sujet, et de ses réponses contrefactuelles.

est évidemment modelée sur la structure des croyances (degrés de croyance) et des désirs (valeurs de préférence) posée par la théorie de la décision⁶², il semble concevoir l'État Mental de *faire comme si* (« make-believe ») comme un État Mental sophistiqué. Or, quoi qu'il vise, ce qu'il offre ne pourrait pas être un État Mental sophistiqué. Tout d'abord, si le bloc d'attribution de l'État Mental de *faire comme si* était saturé, alors ce bloc comporterait (toutes) les croyances usuelles de l'agent aussi bien que (tous) les rapports entre ses croyances et ses *faire comme si* (*make-beliefs*). Or, le rapport entre les *faire comme si* et les croyances usuelles est particulièrement difficile et épineux, n'étant pas compris de manière adéquate, encore moins de manière rigoureuse, à l'heure actuelle⁶³, même dans les théories proposées par Currie et Walton. En conséquence, le bloc d'attribution de l'État Mental de *faire comme si* ne peut pas être saturé (à moins que les rapports entre les États Mentaux attribués dans un même bloc ne soient pas compris, et *a fortiori* ne soient pas compris de manière rigoureuse à la façon de la théorie de la décision). Un raisonnement semblable s'applique relativement à la base d'attribution : si la base d'attribution était globale, alors y figureraient des expériences et des comportements qui ne sont pas relatifs à la fiction (ou le jeu de *faire comme si*), dont les rapports avec les États Mentaux de *faire comme si* sont peu compris à l'heure actuelle⁶⁴, y compris dans les théories de Currie et de Walton. De la même façon, il en découle que le bloc, quitte à contenir des rapports forts obscurs entre États Mentaux, n'a pas de base globale. *Pace* les intentions de Currie, et éventuellement de Walton, l'État Mental de *faire comme si* qu'ils proposent n'est pas sophistiqué à la façon de la croyance de la théorie de la décision, mais plutôt ordinaire à la façon de la croyance dont on parle quotidiennement.

Par conséquent, on traite désormais les États Mentaux proposés comme spécifiques à la fiction (imagination, *faire comme si*) comme des États Mentaux ordinaires. Ce qui ne nuit nullement à l'argumentation car, comme on le montre dans §1.4.2, les caractéristiques des États Mentaux spécifiques qui les empêchent de répondre pleinement à la question du rapport détaillé entre fic-

⁶² « Nous calculons le désir sur l'arrière-fond des possibilités offertes par les croyances de l'agent. Nous calculons le « désir dans le jeu de *faire comme si* » (*make-desire*) sur l'arrière-fond des possibilités pertinentes dans le jeu de *faire comme si* du lecteur » Currie (1990, p205).

⁶³ Ce rapport est un des sujets de la question C. de §1.1, laquelle a été reconnue comme particulièrement épineuse – et non résolue – dans §1.2.1 (voir la discussion du versant sémantique de cette question).

⁶⁴ Puisque les croyances usuelles ont affaire à ces comportements « hors » le jeu de *faire comme si*, la difficulté du rapport est en quelque sens liée à la difficulté du rapport entre le *faire comme si* et la croyance usuelle.

tion et réalité se reproduisent dans les États Mentaux usuels, et notamment les croyances, *qu'ils soient conçus comme ordinaires ou sophistiqués*. Les critiques qui seront adressées aux États Mentaux ne sont pas spécifiques à la conception que l'on leur fait, encore qu'elles prennent des formes légèrement différentes pour les différentes conceptions.

Jusqu'à présent, alors que les notions introduites sont totalement générales, l'exemple considéré est un exemple cognitif, à savoir celui de la croyance. En revanche, le questionnement posé à l'égard de la fiction (§1.1) a trait aussi bien au côté affectif qu'au côté cognitif. Pour permettre une réflexion approfondie sur cette question dans la suite, ainsi que la compréhension des théories des affects particuliers à la fiction proposées par Walton et Currie, il est nécessaire de présenter quelques faits de base ayant trait aux affects, lesquels sont assez largement admis, y compris notamment par Walton et Currie. Pour ce faire, on aura besoin d'une distinction entre deux « *types* » d'État Mental, à savoir entre ceux qui peuvent être attribués faisant référence à un instant seul et ceux qui, pour être attribué à un agent, exigent une référence à plusieurs instants. On expose cette distinction dans le prochain paragraphe, pour en venir à la question des affects dans le paragraphe suivant.

1.3.3 États mentaux narratifs et non narratifs

Où l'on présente la distinction entre les États Mentaux narratifs, tels que la croyance, dont l'attribution requiert une base d'une certaine étendue temporelle et les États Mentaux non narratifs, dont le sentiment, qui peuvent être attribués en faisant référence à un seul instant. Bien entendu, le classement des États Mentaux comme narratifs ou non est relatif aux propriétés de la base d'attribution admise : pour certains choix de base, dont le celui pris par l'externalisme méthodologique le plus strict, il n'y aura pas d'États Mentaux non narratifs.

Dans le paragraphe précédent, on a mis en évidence une distinction entre deux *conceptions* des États Mentaux, entre deux *conceptions* de la croyance par exemple, dont une des différences tient à la base de l'attribution. Il s'agit ici d'introduire une distinction entre deux types d'*États Mentaux particuliers*, entre la croyance et le sentiment par exemple, qui s'exprime également en termes de la base de l'attribution. En l'occurrence, il est question de la *possibilité* d'une attribution sur une base plus ou moins grande ou petite : un État Mental particulier *peut-il* être attribué en faisant référence à un instant seul ou toute attribution

doit-elle faire référence à d'autres moments ? Cette question, et la distinction qui sera tirée selon la réponse, n'est pas triviale, dans la mesure où il y a à première vue des États Mentaux pour lesquels la réponse est négative et d'autres pour lesquels elle est apparemment positive.

D'un côté, on s'accorde généralement sur le fait qu'une croyance est en rapport avec des expériences du sujet (qui y donnent lieu) et avec des actions du sujet (qui sont conséquences des croyances, et éventuellement des désirs) à d'autres instants⁶⁵. On dira que la croyance est un État Mental *narratif* : son attribution dépend nécessairement non seulement de l'état instantané du sujet, mais plutôt d'un certain nombre de ses expériences et de ses comportements, parsemés dans le temps et organisés sous forme d'un récit. De l'autre côté, un sentiment est apparemment quelque chose de « présent », de « vif » ou « d'immédiat » (*occurrent*), donc relatif à un instant seul, et donc admettant en principe d'être attribué sur la base de ce seul instant. En appelant *non narratif* un État Mental dont l'attribution peut se faire en référence à un instant seul, le sentiment est apparemment un exemple d'un État Mental non narratif.

Cette distinction entre les États Mentaux narratifs et non narratifs est entièrement rigoureuse et générale, encore qu'elle soit bien entendu relative à ce que l'on prend comme base d'attribution. Si l'on prend une base d'attribution munie des propriétés particulières, la distinction s'avérera vide. Par exemple, les seuls États Mentaux que l'on peut admettre comme non narratifs sont généralement ceux qui sont phénoménologiquement « présents » ou « immédiats » (*occurrent*), de telle sorte que dans un cadre théorique souscrivant à l'externalisme méthodologique le plus stricte (§1.3.2), n'admettant aucune référence aux expériences phénoménales du sujet, la distinction entre États Mentaux narratifs et non narratifs n'a pas de sens. Dans un tel cadre souscrivant à une forme stricte d'externalisme méthodologique, le sentiment, exemple typique d'un État Mental non narratif, doit être attribué en s'appuyant sur des comportements étalés temporellement⁶⁶.

⁶⁵Par exemple, Ryle, *op cit.*, p120 soutient que les affirmations des dispositions, dont, à ses yeux, les attributions sont « étroitement liées avec les récits d'événements, car, si elles sont correctes, elles sont satisfaites par les récits d'événements ». Ou encore, Stalnaker (1984, p19) affirme que « Les croyances ont un contenu déterminé en raison de leurs connexions causales supposées avec le monde. Les croyances sont *croyances*, plutôt que quelque autre état représentationnel, en raison de leur rapport ... avec l'action. »

⁶⁶Wittgenstein (1953, §244) : « Comment les mots *se rapportent-ils* aux sensations ? ... Cette question est semblable à : comment un homme apprend-il la signification des noms des sensations ? – du mot « douleur » par exemple. En voici une possibilité : des mots sont liés aux expressions primitives, naturelles de la sensation et employés à sa place. Un enfant s'est blessé, il crie ; alors les adultes lui parlent, et lui apprennent des exclamations et, plus tard,

Il ne reste pas moins que la distinction entre États Mentaux narratifs et non narratifs est parfois explicitement utilisée, et souvent effectivement importante. Elle permet par exemple d'expliciter une propriété importante du *faire comme si* proposé par Walton et Currie, à savoir qu'il est un Etat Mental *narratif*. En effet, aussi bien Walton que Currie soulignent le caractère éventuellement non immédiat (*nonoccurrent*) du *faire comme si* : « je ne pense pas le *faire comme si* comme un état « qualitatif » ou « phénoménologique », ouvert à l'introspection à la façon des douleurs et des sensations corporelles » ⁶⁷. Il est sous-entendu que, puisque le *faire comme si* ne se réduit pas à ses apparitions « phénoménologiques », il est attribué sur la base des comportements et des expériences relatifs à d'autres instants temporels : en un mot, le *faire comme si* est, tout comme la croyance, narratif. Il en découle la conséquence importante que le *faire comme si*, tout comme la croyance, pourrait être attribué à un sujet sans qu'il se rende compte (explicitement) qu'il ait cet État Mental.

En outre, comme on le verra tout de suite, cette distinction est cruciale pour comprendre le débat autour des émotions, et particulièrement les positions de Currie et de Walton à l'égard des affects relatifs à la fiction.

1.3.4 Émotions et sentiments

Dans le champ affectif, il y a une distinction importante et répandue entre les sentiments, qui sont immédiatement présents et par conséquent non narratifs, et les émotions, qui sont des États Mentaux narratifs ayant des liens forts et constitutifs avec plusieurs comportements, croyances, désirs, attitudes et sentiments. Cette distinction permet de comprendre les oppositions entre les théories proposées des affects relatifs à la fiction : elles s'accordent généralement sur la question des sentiments à l'œuvre dans la réception de la fiction, et de leurs rapports avec d'autres attitudes, mais divergent dans le classement des émotions constituées par ces sentiments et ces rapports.

Le questionnement formulé à l'égard de la fiction dans §1.1 a trait non seulement à ses aspects cognitifs, mais également à ses aspects affectifs (notamment la question **B.**). Comme ailleurs dans ce travail, il s'agit non pas de la question de ce *qu'est* une émotion (éprouvée relativement à la fiction), mais plutôt la question *des effets et des apparitions* des phénomènes affectifs qui ont trait à la

des phrases. » Voir également Bouveresse (1976, Ch. 4, notamment §4.)

⁶⁷Currie (1990, p21). Walton (1990, p16-18) va dans le même sens.

fiction, bref, de *la structure* des émotions et de *leurs rapports* avec des comportements, des phénomènes, et d'autres États Mentaux. En conséquence, l'intérêt des débats autour de la question de l'émotion pour le propos de cet essai se borne aux *structures* proposées, et ne s'étend pas aux questions métaphysiques ou ontologiques générales qui peuvent être formulées à leur égard. Il s'avère que, en raison du manque général de discussion sur les éventuels mécanismes affectifs relatifs à la fiction, on pourra faire l'économie de la question « causale » des émotions (quels mécanismes produisent ou constituent des émotions ?), pour concentrer sur la question « conceptuelle » des émotions : quelles sont les modalités d'attribution de l'émotion ?

L'enjeu de la discussion suivante, portant sur la question conceptuelle des émotions, est double. D'une part, elle permet de comprendre les théories de Walton et de Currie ayant trait aux affects relatifs à la fiction. En parcourant quelques théories générales des affects, on parvient à distinguer le *sentiment*, aspect affectif « phénoménologiquement présent » et non narratif, et *l'émotion*, aspect affectif narratif, rapporté à d'autres États Mentaux, attitudes, comportements et éventuellement sentiments du sujet. Il s'avérera que les théories de Walton et de Currie sont d'accord pour ce qui concerne les sentiments impliqués dans la lecture de la fiction, aussi bien que sur le rapport entre ces sentiments et d'autres États Mentaux, mais qu'elles diffèrent dans leur catégorisation de l'État narratif rapportant au sentiment et à ces États Mentaux : Currie l'appelle émotion, Walton l'appelle « émotion imaginée ».

D'autre part, et de plus, la discussion mettra en évidence quelques éventuelles exigences sur une théorie psychologique de la fiction qui vise également son versant affectif. En effet, sont repérés ici quelques aspects spécifiques des émotions qui *pourraient* être exigés d'une telle théorie, dont certains, notamment la nécessité d'un Etat Mental non narratif et phénoménologiquement présent de sentiment, sont débattus. Dans les chapitres suivants, sera développée une théorie psychologique générale de la fiction qui contient une théorie du sentiment ; en conséquence, cette théorie *peut* rendre compte des émotions, que les sentiments soient des aspects vraiment requis ou non.

L'aspect spécifique aux affects qui est souvent promu dans la littérature est le *sentiment*. Le sentiment est quelque chose qui *a lieu*, il « *fait occurrence* ». Il est quelque chose de « *présent* », de « *vif* » et « *d'immédiat* »⁶⁸ ; plus précisément, il est « *présent* », « *vif* » et « *immédiat* » *phénoménologiquement*⁶⁹. Les sentiments

⁶⁸Pourtant, la « présence » du sentiment n'implique pas sa « luminosité » : on pourrait avoir un certain sentiment sans tenir compte qu'on l'éprouve.

⁶⁹Bien entendu, on fait abstraction des complexités impliquées par l'adoption d'un exter-

prennent une place importante dans un grand nombre de théories de l'émotion. Le cas le plus extrême est celui de James, selon lequel « notre sentiment de ces mêmes changements [corporels] en tant qu'ils se produisent *est* l'émotion »⁷⁰. Or, on trouve un accent sur l'importance des sentiments également chez Elster, selon lequel un des traits principaux des émotions sont leurs « sentiments qualitatifs », ou chez Goldie, selon lequel « les sentiments constituent une partie intime et familière des expériences émotionnelles »⁷¹, pour ne prendre que deux exemples.

Par contraste, un grand nombre de philosophes soulignent les rapports des émotions avec les comportements du sujet, avec d'autres États Mentaux tels que les croyances et les désirs, ou avec des attitudes tels que les jugements. Certains prétendent que ces rapports sont les seuls qui importent quant aux émotions, ainsi faisant fi des éventuels sentiments. Par exemple, Bedford affirme que « les termes d'émotions ... présentent l'action qu'il s'agit d'expliquer non seulement dans le contexte du comportement de l'individu, mais dans un contexte social, ... [et ils] expliquent en fournissant une raison pour l'action, c'est-à-dire en en fournissant une justification compétente ou partielle » ; selon Solomon, « mes émotions *sont* les jugements que je fais » ; ou encore selon Davidson « en fournissant la croyance [que l'on a une certaine caractéristique] et l'attitude [d'approbation ou d'estime pour toute personne qui a cette caractéristique] sur lesquelles l'orgueil repose, on explique l'orgueil dans deux façons : on en fournit une explication causale et on fournit les raisons de la personne pour son orgueil »⁷². Même ceux qui n'entendent pas réduire les émotions aux autres États Mentaux, aux attitudes, ou aux comportements seulement, reconnaissent généralement ces aspects. Elster, par exemple, compte parmi les caractéristiques principales des émotions leurs « antécédents cognitifs », et Goldie concède que « lorsque j'ai une émotion, il y aura souvent des croyances et des désirs qui peuvent être m'attribués et qui contribueront à rendre intelligible aussi bien mon émotion que ce que je fais en conséquence de cette émotion »⁷³.

Cet bref aperçu de la littérature permet deux conclusions. Premièrement, de même que les croyances, les émotions sont des États Mentaux *narratifs*, en ce sens que leur attribution dépend de plusieurs expériences et comportements à des moments différents, aussi bien d'ailleurs que d'autres États Mentaux attribués tels les croyances et les désirs. Comme le dit très clairement Goldie,

nalisme méthodologique ; voir le paragraphe précédent.

⁷⁰James (1884, pp189-90).

⁷¹Elster (1999a, p 246 *sq.*) ; Goldie (2000, p50).

⁷²Bedford (1957, p97-8) ; Solomon (1973, p261) ; Davidson (1976, pp284-5).

⁷³Elster, *ibid.* ; Goldie, *op cit.*, p18.

« une émotion est structurée de sorte qu'elle constitue une partie d'un récit – en gros, une suite des actions et des événements, des pensées et des sentiments – dans lequel l'émotion s'enfonce »⁷⁴. La jalousie, pour reprendre son exemple, est en rapport avec un certain nombre de croyances (la femme aimée vous a quitté pour quelqu'un d'autre), avec un certain nombre de comportements agités (le manque de sommeil), avec un certain nombre de pensées et de sentiments (relatifs à cette femme), et avec un certain nombre d'actions (vous la cherchez) qui sont parsemés dans le temps et organisés sous forme de récit. Il découle de cette « narrativité » une conséquence déjà remarquée dans les cas de croyance et de *faire comme si* (§1.3.2), à savoir le fait qu'une émotion pourrait être attribuée à quelqu'un à juste titre sans qu'il soit conscient de cette attribution : ainsi, la narrativité de l'émotion se joint au manque bien connu de limpidité du sujet à l'égard de ses propres émotions (et d'ailleurs à son défaut d'« autorité » relativement à ses émotions)⁷⁵.

Pourtant, et voilà la deuxième conclusion, à la différence des croyances qui sont apparemment attribuées sur la base des comportements du sujet, il *se peut* que l'émotion requière quelque chose de plus que les comportements, à savoir quelque chose de l'ordre du *sentiment*. Comme suggéré par les citations ci-dessus, ceci est une question débattue ; néanmoins, *si* l'on propose une théorie qui *peut* rendre compte *aussi bien* des sentiments *que* des croyances et des désirs, cette théorie *sera* capable de rendre compte également des émotions (qu'il faille employer les sentiments ou non)⁷⁶. La théorie proposée dans les Chapitres 2 et 3 aura cette propriété.

Dans la suite, le terme *sentiment* sera réservé pour ces affects immédiats et présents que l'on éprouve « dans le moment ». Par contraste, le terme *émotion* réfère à l'Etat Mental attribué à un sujet sur une base comportant plusieurs moments différents ; comme telle, elle « comporte plusieurs facettes », relatives aux différents comportements, États Mentaux, sentiments et ainsi de suite qui y sont en rapport. Ces rapports sont établis dans le bloc d'attribution (§1.3.2) où figure l'émotion particulière attribuée : les autres États Mentaux (tels les croyances et les désirs) auxquels cette émotion se rapporte sont également présents dans

⁷⁴*Op cit.*, p13.

⁷⁵Comme le dit Bedford, *op cit.*, p81 « ceux qui sont jaloux sont souvent les derniers, plutôt que les premiers, à le reconnaître. » Solomon, *op cit.*, pp259-60, pour prendre un autre exemple au hasard, suggère avec Freud que certaines émotions sont « dissoutes » dès que le sujet en prend proprement conscience.

⁷⁶Il se peut que l'attribution des émotions s'appuie également sur quelque chose de l'ordre d'une *pensée*, conçue comme un aspect cognitif « présent » et « immédiat » (*occurrent*) : une sorte d'analogie cognitive pour le sentiment. Il s'avère que la théorie proposée rendra également compte de telles pensées ; voir §2.5.1.

ce bloc. Il est clair des propriétés de ces blocs que la conception quotidienne d'émotion, ainsi que celle que l'on emploie dans le milieu philosophique, est une conception *ordinaire*. Intuitivement, on n'a que des formulations génériques et non rigoureuses au sujet des rapports entre les émotions et d'autres États Mentaux (cf. la citation de Davidson ci-dessus, par exemple), ressemblant plutôt aux sagesses quotidiennes vagues qu'aux principes précis de la théorie de la décision. Plus précisément, la conception d'émotion employée est *non saturée* et a une *base d'attribution locale* : d'un côté, il n'y a aucune tentative de mettre une émotion en rapport avec *tous* les autres États Mentaux du sujet ; de l'autre côté, les discussions philosophiques à l'égard des émotions ont trait normalement aux cas particuliers, supposés génériques, plutôt que sur la vie entière du sujet.

Cette conception générale d'émotion est, *en gros*, admise par Walton et par Currie et employée dans leurs considérations des affects relatifs à la fiction. Les deux auteurs admettent que, tout comme dans la réalité, on ressent certains *sentiments* en réponse à un texte fictionnel (ou un film fictionnel)⁷⁷. En outre, ils s'accordent à ce que la différence entre les émotions usuelles et les « émotions » relatives à la fiction tient au fait que, alors que les premières entrent en relation avec certaines croyances, désirs (et enfin actions) *usuels*, les dernières entretiennent le même genre de relation avec les États Mentaux de *faire comme si* (et, selon Currie, de « make-desire »)⁷⁸ : la peur du spectateur du cinéma est issue de son « *faire comme si* » que le héros est en danger, non de quelque croyance à cet égard. Or, ils s'opposent sur la question de la priorité à accorder à cette différence relativement à la ressemblance entre émotions usuelles et « fictionnelles » (il s'agit dans les deux cas du même genre de rapport avec des États Mentaux cognitifs et volitifs). Currie minimise la différence au profit de la ressemblance entre elles, en affirmant que il s'agit bel et bien des émotions dans la fiction, mais des « espèces différentes d'émotion »⁷⁹. En revanche, Walton favorise la différence par rapport à la ressemblance : le spectateur d'un film d'horreur, pour reprendre son exemple, *imagine* qu'il est en danger et donc il *imagine* avoir peur. Autrement dit, puisque le danger est fictionnel, l'émotion est également « fictionnelle »⁸⁰. Ainsi, l'opposition entre les deux auteurs ressemble à une différence de terminologie plutôt qu'une différence de fond ; pour

⁷⁷Currie appelle ces sentiments des sentiments (*op cit.*, §5.3-4), alors que la conception waltonienne de « quasi-émotion » (*op cit.*, p196 et §7.1) correspond *grosso modo* à ce que l'on appelle ici sentiment.

⁷⁸Currie, *op. cit.*, §§5.4 et 5.5 ; Walton, *op cit.*, §7.1.

⁷⁹Currie, *op. cit.*, p212.

⁸⁰Walton, *op cit.*, p247.

les propos de la discussion suivante, les deux théories sont logées à la même enseigne.

1.4 Approches « conceptuelles » : États Mentaux spécifiques à la fiction

On a proposé deux genres de réponse à la question du rapport entre fiction et réalité dans la psychologie du lecteur : l'une consiste à élaborer des conceptions qui permettent de comprendre les comportements relatifs à la fiction, c'est-à-dire des *États Mentaux* à attribuer au lecteur de la fiction ; l'autre prétend « expliquer » ces comportements en proposant des *mécanismes psychologiques* qui leur sont « sous-jacents ». De plus, toutes les théories proposées de la psychologie de la fiction jusqu'à cette date adoptent la stratégie « mimétique », consistant à poser la distinction cruciale entre la fiction et la réalité, pour ensuite retrouver d'une manière ou d'une autre la similitude entre elles (§1.2.1). Toutes ces approches, qu'elles s'occupent de la question des États Mentaux ou de celle des mécanismes, se heurtent, *en gros*, à trois difficultés qui les empêchent de pouvoir répondre pleinement à la question de la fiction posée dans §1.1. En premier lieu, le rapport entre l'aspect psychologique proposé spécifiquement pour la fiction (État Mental, mécanisme) et le reste de la psychologie du lecteur (États Mentaux ou mécanismes « usuels ») est peu compris, surtout à certains moments importants (moments d'immersion fictionnelle, d'entrée ou de sortie de la lecture, par exemple). En second lieu, le rapport entre l'aspect psychologique proposé spécifiquement pour la fiction (État Mental, mécanisme), les aspects psychologiques ordinaires du lecteur (États Mentaux ou mécanismes « usuels ») et l'expérience phénoménologique du lecteur est peu compris, notamment à ces mêmes moments. En dernier lieu, la distinction rigide qu'elles posent entre un aspect psychologique spécifique à la fiction (État Mental, mécanisme) et les aspects psychologiques usuels supporte mal les apparentes modifications dans les modalités fictionnelles entre des lectures, des œuvres, des genres littéraires ou éventuellement des époques différents.

Or, les théories des États Mentaux *usuels* et des mécanismes psychologiques *usuels* rencontrent des difficultés tout à fait semblables. Ceci est important en raison du soutien supplémentaire apporté à l'approche qui sera finalement empruntée dans les chapitres suivants. Car, comme on l'a expliqué dans §1.2.3, cette approche ne suppose *ni* les États Mentaux spécifiques à la fiction *ni* les États Mentaux usuels (respectivement, *ni* les mécanismes psychologiques spécifiques à la fiction *ni* les mécanismes psychologiques ordinaires), mais prend

son départ d'un « cadre » ou d'un « niveau » psychologique où la distinction entre fiction et réalité n'est pas donnée ni apparente *a priori*. La persistance des faiblesses touchant aux États Mentaux (mécanismes) spécifiques dans le cas des États Mentaux (mécanismes) usuels met en doute la stratégie d'aborder la question détaillée relative à la fiction dans ce(s) cadres psychologiques, et sert par conséquent à motiver le recours à un autre cadre psychologique. En outre, l'analyse de ces faiblesses qui sera entreprise dans cette section et la prochaine sera utile à la constitution de cet autre cadre.

Bien que les difficultés se posent de manière générale, il convient de les élaborer en deux grands moments : dans cette section, il s'agit des États Mentaux, dans la section suivante, il s'agira des mécanismes. Chacun de ces moments se répartit en deux : le premier paragraphe ici porte sur les États Mentaux spécifiques à la fiction, dans le deuxième paragraphe, on montre comment des difficultés semblables se posent pour d'autres États Mentaux usuels.

1.4.1 Difficultés pour la théorie des États Mentaux spécifiques à la fiction

Où l'on met en évidence trois difficultés que rencontre la théorie des États Mentaux spécifiques à la fiction. Tout d'abord, le rapport entre les États Mentaux relatifs à la fiction et les États Mentaux usuels à certains moments importants est beaucoup plus complexe que ne le suggère l'opposition simple posée par la théorie des États Mentaux spécifiques à la fiction. Ensuite, alors que l'expérience phénoménale du lecteur à ces moments ne correspond pas à la simplicité des États Mentaux qui lui sont attribués, la théorie des États Mentaux spécifiques à la fiction n'explique pas ce décalage. Enfin, il y a apparemment une tension entre la relative stabilité de l'État Mental spécifique à la fiction posé par cette théorie et la variabilité de ses modalités selon la lecture, le genre ou l'époque.

La théorie des États Mentaux spécifiques à la fiction proposée par Currie et Walton (§1.2.1) parvient à concilier la nécessaire *distinction* entre les États Mentaux usuels (*croire que Holmes n'existe pas*) et les États Mentaux impliqués dans la lecture et la compréhension de la fiction (*faire comme s'il existait*) et leur non moins nécessaire *ressemblance* (*faire comme si* ressemble à *croire*). En revanche, si l'on pousse plus loin le questionnement au sujet de la fiction, pour s'intéresser aux détails plus précis du rapport entre les États Mentaux spécifiques à la fiction proposés et les États Mentaux usuels, cette théorie rencontre trois

difficultés.

1. Il y a apparemment une relation complexe entre les États Mentaux relatifs à la fiction et les États Mentaux usuels à *des moments particuliers* et *importants*, tels les moments de pleine immersion dans la lecture, les moments où la lecture est interrompue, ou les moments où les faits réels entrent en jeu, laquelle est peu compréhensible en termes de la simple dichotomie entre États Mentaux spécifiques à la fiction (*faire comme si*) et États Mentaux usuels (croyances).
2. À ces moments cruciaux, l'expérience phénoménale immédiate du lecteur de fiction (ce qu'il *éprouve*) ne correspond tout simplement pas à la simple dichotomie entre les États Mentaux spécifiques à la fiction et les États Mentaux usuels.
3. Contrairement aux hypothèses de stabilité trans-générique, trans-interprétative et trans-historique implicitement adoptées par les théoriciens des États Mentaux spécifiques, il y a apparemment la *possibilité* de changements des modalités du *faire comme si* selon les textes, les lectures et peut-être les époques : le *faire comme si* impliqué dans la lecture d'un roman policier n'est pas exactement celui qui convient à la lecture d'un roman post-moderne, celui qui a cours aujourd'hui diffère probablement de celui qui était en vigueur au XVII^e siècle, celui qui est mobilisé dans une interprétation marxiste du *Procès* de Kafka ne ressemble que très peu celui qui sous-tend une lecture psychanalytique.

Rapports entre États Mentaux usuels et spécifiques à la fiction à des moments particuliers

Pour mettre en évidence la première difficulté, on considère trois moments où la relation entre les États Mentaux cognitifs spécifiques à la fiction et les États Mentaux cognitifs usuels affiche des propriétés apparemment manquées par les théories proposées, et un moment où un problème semblable se pose dans le domaine affectif. Dans tous les cas, le problème consiste en le fait que la complexité de la relation entre les États Mentaux spécifiques à la fiction et les États Mentaux usuels ne correspondent point à la simplicité de l'opposition posée par Walton et Currie.

Un premier exemple concerne le rapport entre les États Mentaux spécifiques et les États Mentaux usuels à l'égard de l'existence de Sherlock Holmes (question A. de §1.1.1) à un instant de la lecture. À un tel instant, le lecteur se dit croire que Sherlock Holmes existe, ce qui est, selon Currie et Walton, une

locution de l'ordre du *faire comme si*. Or, au même moment où il se dit (dans le « *faire comme si* ») croire que Holmes existe, il croit (authentiquement) qu'il n'existe pas : mais alors, comment comprendre la valorisation apparente du *faire comme si* sur la croyance ? Les théories de Currie et de Walton n'offrent aucune compréhension des particularités de ce rapport entre le *faire comme si* et la croyance à de tels instants. Mais ce rapport est *crucial* pour comprendre la maîtrise de la distinction entre fiction et réalité par le lecteur, son évitement de la confusion entre les deux et la possibilité même de son appréciation de la fiction, bref, une grande part de ce que l'on désigne, depuis Schaeffer, sous le nom de la « compétence fictionnelle » du lecteur (§1.1.3). Comment comprendre son appréciation dans la lecture et l'absence de confusion avec la réalité, si l'on ne comprend pas *au moins* le rapport entre l'inexistence (réelle) et l'existence (fictionnelle) de Holmes en tant qu'elles prennent place dans sa psychologie à un moment donné de lecture, que ce soit comme États Mentaux différents ou autrement ? Et pour comprendre le fait que la croyance à l'inexistence de Holmes ne nuise pas à « l'immersion fictionnelle » du lecteur (§1.1.3), ne faut-il pas au moins dire quelque chose sur la *priorité* accordée à son *faire comme si* (relatif à l'existence de Holmes) à ce moment ? Les théories de Currie et de Walton ne se prononcent pas sur ce point.

Un autre cas qui témoigne de ce genre de défaut des théories de *faire comme si* a trait à un moment où le jeu de *faire comme si* est rompu, par exemple si on interrompt le lecteur pour lui poser la question de l'existence de Holmes. Pour répondre à la question, le lecteur doit résoudre le conflit apparent entre sa croyance à l'inexistence de Holmes, et son *faire comme si* il existait, en choisissant de répondre dans le registre du *faire comme si* (*il existe*) ou dans le registre réel (*il n'existe pas*). Pourtant, à un tel moment d'hésitation avant sa décision, ces États Mentaux entrent dans des rapports complexes, dont la simple opposition posée par Walton et Currie ne parvient pas à rendre compte. Néanmoins, tels rapports à tels moments sont tout à fait capitaux pour comprendre les modalités d'entrée et de sortie du jeu de *faire comme si*, question sur laquelle les théories du *faire comme si* gardent le silence.

Ce problème de l'écart entre l'opposition simple de Currie et Walton et la relation complexe effective entre États Mentaux usuels et spécifiques à des moments particuliers se pose également relativement aux faits réels pertinents et importants pour la lecture de certains passages du texte. Pour reprendre l'exemple de la croyance « usuelle » que les humains ne survivent pas à des chutes du haut des escarpements élevés, laquelle est nécessaire pour conclure, lors d'un passage où un personnage est dans cette position, qu'il est en danger

(question **C.** de §1.1.1), il paraît que, au moment approprié de la lecture, le rapport entre cette croyance usuelle et l'État Mental spécifique de *faire comme si* devient plus complexe : cette croyance, pour ainsi dire, *entre en scène*. Or, du point de vue de l'attributeur, le lecteur « a » cette croyance tout au long de sa lecture, d'où une difficulté de comprendre le changement qui se produit lorsque cette croyance *entre en scène*, c'est-à-dire lorsqu'il « entre dans » le jeu de *faire comme si*. Les théories proposées ne sont pas capables de clarifier le changement dans l'« activité » de la croyance qui correspond intuitivement à son entrée en scène.

Le problème relevé, illustré jusqu'à ici sur les exemples relatifs aux États Mentaux cognitifs (*faire comme si*, croyance), se pose de manière semblable sur le plan affectif (question **B.** de §1.1.1). Dans le cas affectif, Walton et Currie n'en viennent qu'à concevoir le rapport entre les affects relatifs à la fiction et les États Mentaux cognitifs spécifiques à la fiction (*faire comme si* ou « imagination »), ce qui n'est pas suffisant pour penser les affects qui entrent en relation aussi bien avec les États Mentaux cognitifs spécifiques à la fiction (*faire comme si*) qu'avec les États Mentaux cognitifs usuels (croyance). Lorsque le spectateur de cinéma – paralysé de peur – se répète que « ce n'est qu'une fiction », il semble y avoir une tension entre une croyance (réelle), les affects du lecteur, et les aspects de son *faire comme si*. Le *faire comme si* relatif à la situation présentée sur l'écran incite les affects de peur, grâce aux rapports entre les *faire comme si* (*make-beliefs*, « croyances imaginées ») et les affects ; par contraste, la croyance au caractère fictionnel de la situation devrait neutraliser ces affects, grâce aux rapports entre les croyances (réelles) et les affects. De même que dans le cas cognitif, il n'y a chez Walton et Currie aucune compréhension de ces situations et de ces rapports complexes. Alors qu'en général, la discussion de cette section et de la prochaine s'applique sur le plan affectif, on se concentre désormais sur le cas cognitif, qui est plus développé dans la littérature.

En somme, il y a un certain nombre de moments où le rapport effectif entre États Mentaux relatifs à la fiction et États Mentaux habituels est apparemment beaucoup plus complexe que n'est l'opposition simple posée par Currie et Walton. Pourtant, la portée de cette difficulté étend au-delà de ces simples moments particuliers : pour dire vite, la propre compréhension du rapport entre les États Mentaux à des moments particuliers est une *condition préalable* à toute compréhension de la *dynamique* de ce rapport. Dans la mesure où la lecture est un processus *dynamique*, où les croyances réelles *entrent* en jeu (et en *sortent*), où les comportements du lecteur à l'égard du texte (ses affirmations, ses conduites affectives) *se déroulent*, où tous ces aspects mettent en œuvre des

configurations différentes des États Mentaux relatifs à la fiction et États Mentaux habituels, la question de la dynamique du rapport entre États Mentaux spécifiques à la fiction et États Mentaux usuels touche au cœur du rapport entre réalité et fiction dans la psychologie du lecteur. Comment comprendre la persistance de l'immersion fictionnelle si l'on ne comprend pas la relative *absence* des croyances authentiques qui risquerait d'y nuire ? Comment comprendre la réponse du lecteur à la question de l'existence de Holmes posée pendant la lecture si l'on ne comprend pas les complexités de sa délibération ? Comment comprendre *quelles* croyances entrent en jeu dans la lecture, et leurs conséquences, si l'on ne comprend pas comment une croyance « entre » dans le jeu de *faire comme si* ? En outre, puisque ces moments problématiques sont des moments *clefs* dans des démarches importantes de la lecture et de l'appréciation de la fiction (l'immersion fictionnelle, l'entrée en scène de croyances authentiques, l'entrée et la sortie du jeu de *faire comme si*), ces démarches mêmes échappent elles aussi aux théories des États Mentaux spécifiques à la fiction. Par conséquent ces théories échouent à aborder pleinement la question de la fiction et de son rapport à la réalité (§1.1).

Expériences phénoménales du lecteur et ses États Mentaux spécifiques à la fiction

Alors que la première difficulté a trait à la *structure des rapports entre États Mentaux*, la seconde difficulté concerne au contraire la *phénoménologie du lecteur*⁸¹ : à certains moments, il y a un décalage entre l'expérience phénoménale du lecteur et les États Mentaux que l'on lui attribue. En vérité, les moments où se pose la première difficulté sont normalement les moments où se pose la seconde, de sorte que les mêmes exemples sont pertinents. En considérant d'abord un moment de la lecture, le fait que Holmes n'existe pas (une croyance du lecteur) ne figure pas dans la « vie mentale immédiate » du lecteur, à moins que ce soit dans un arrière-plan extrême, par opposition à son existence (un *faire comme si*) qui est constamment en jeu, quoique tacitement, en conséquence du fait que le lecteur lui attribue des actions et des attitudes. La relation entre les États Mentaux posée par Currie et Walton n'affiche aucune telle différence de *présence*. Passant ensuite au moment où l'on pose au lecteur la question de l'existence d'Holmes, le lecteur a, dans sa « vue de l'esprit », aussi bien l'existence « fictionnelle » de Holmes que son inexistence effective, enchevêtrées

⁸¹ Ainsi, alors que l'on peut nier la seconde difficulté en appliquant les contraintes de l'externalisme méthodologique (§1.3.2), la première difficulté « a toujours corps » dans l'optique externaliste.

apparemment dans un rapport complexe. Cette complexité n'apparaît pas au niveau des États Mentaux proposés. Revenant enfin au moment où la croyance relative aux chutes nuisibles à la santé entre en scène, cette nuisibilité est associée à l'expérience phénoménale à ce moment, si ce n'est que « au bord » de cette expérience, seulement « tacitement » ou « implicitement », car, sans *l'activité* du fait de la nuisibilité de telles chutes dans l'expérience phénoménale, le lecteur n'éprouverait nullement d'inquiétude à l'égard du personnage. Cette « mise en activité » n'est cependant pas compréhensible dans les théories de Currie et de Walton.

Cette seconde difficulté est apparemment reliée à la première ; plus précisément, il semble que, à un moment donné, les relations entre les éléments apparaissant dans l'expérience phénoménologique du lecteur correspondent aux relations entre les États Mentaux du lecteur à l'égard de ces éléments. Par exemple, l'*expérience phénoménale* à un moment d'immersion fictionnelle (l'existence de Holmes est en jeu, mais non pas son inexistence crue) semble refléter la priorité du *faire comme si* (relatif à l'existence de Holmes) sur la croyance (à son inexistence). Encore, la complexité du rapport entre l'existence « fictionnelle » et l'inexistence effective de Holmes en tant qu'elles apparaissent dans la phénoménologie du lecteur au moment où il est questionné ressemble à la complexité du rapport entre les deux États Mentaux (*faire comme s'il existait* et croyance à son inexistence), remarquée ci-dessus. Pour reprendre le dernier exemple, « l'activité phénoménologique » d'un fait réel au moment de la lecture où il entre en jeu s'apparente à « l'activité théorique » de la croyance relative à ce fait, qui a été mise en évidence plus haut.

Pour mettre en évidence la première difficulté à laquelle se heurtent les théories de Currie et de Walton, on s'est concentré sur les moments particuliers où les rapports entre les États Mentaux étaient trop complexes pour les théories ; pourtant, la difficulté elle-même s'étend au-delà de ces seuls moments, dans la mesure où une compréhension des rapports entre États Mentaux à ces moments est requis pour une compréhension de la dynamique de ces rapports. De la même façon, alors que l'on a mis en évidence les moments particuliers où la phénoménologie du lecteur échappe aux théories de Walton et de Currie, il en découle que ces théories ne savent pas non plus comprendre la *dynamique* des expériences phénoménales du lecteur. Encore une fois, dans la mesure où l'expérience phénoménale *reflète* les rapports complexes et difficiles entre États Mentaux, le manque de compréhension des changements de la phénoménologie du lecteur va de paire avec l'échec de comprendre les changements des rapports entre ses États Mentaux.

Variabilité des modalités du *faire comme si*

La troisième difficulté rencontrée par les théories proposant un État Mental de *faire comme si* (ou « d'imagination ») spécifique à la fiction tient au conflit entre l'apparente stabilité du *faire comme si* et l'apparente variabilité des modalités du *faire comme si* selon l'époque, le genre et la lecture. D'un côté, dans la mesure où le sens d'un État Mental, ou, si l'on préfère, son « identité », est relatif au *bloc* par lequel l'État Mental est attribué (§1.3.2) au sujet, et étant donné qu'un tel bloc comprend d'autres États Mentaux, la proposition d'un seul État Mental de *faire comme si* semble supposer qu'il entretient des rapports relativement *fixes* avec les autres États Mentaux attribués en bloc avec lui, dont notamment les croyances. Ces rapports fixes n'impliquent aucunement que les croyances réelles mises en jeu dans la lecture de la fiction sont toujours les mêmes – la lecture d'une histoire fictionnelle qui a lieu à Paris appelle des croyances différentes que la lecture d'une histoire qui a lieu à Londres –, ils impliquent plutôt que *le mode de sélection* des croyances mises en jeu sur la base des propos du texte est *constant*.

En revanche, il paraît que le mode de sélection effectivement à l'œuvre dans les lectures des textes de fiction pourrait *varier*⁸², d'abord selon le texte, ensuite selon la lecture et enfin, peut-être, selon l'époque. Tout d'abord, on constate des différences entre les lectures de deux textes de « types » ou de « genres » différents par un même lecteur : le mode de sélection à l'œuvre dans la lecture de *Casino Royale* de Fleming, qui met en jeu généralement les croyances de base à l'égard du monde, ne ressemble pas beaucoup au mode de sélection convenable à la lecture de *Ulysse* de Joyce, qui doit apparemment renvoyer également à d'autres « sortes de croyances », relatives à l'histoire et à d'autres œuvres littéraires, aux aspects psychologiques ou psychanalytiques, aux dimensions méta-textuelles, et ainsi de suite. La collection de croyances du lecteur ne change pas entre les lectures des deux textes, donc il faut que la manière dont ces croyances sont mises en œuvre, c'est-à-dire leur mode de sélection, diffère entre elles. L'existence des éventuelles différences entre les lectures de textes de « types » ou de « genres » différents est bien reconnue, étant souvent renvoyée à des différences de « codes » employés par le lecteur ou de « contraintes pragmatiques » qu'il respecte⁸³ ; ces « codes » ou ces « contraintes pragmatiques »,

⁸²On laisse de côté la fameuse question de s'il faut ajouter ses propres croyances où les croyances du public au moment de l'écriture du texte ; voir par exemple Walton, *op cit.*, Ch 4.

⁸³Eco (1985) pense ces possibilités de lectures différentes comme ayant trait à quelque chose de l'ordre de la « pragmatique » (§3.2) ; Stempel (1979), tout en soulignant qu'il n'y a pas de « correspondance » entre les « modalités de réception » et les genres (ou ce qu'il appelle les

dans la mesure où ils mettent en rapport les propos du texte et les croyances authentiques du lecteur, ressemblent à ce que l'on appelle ici le mode de sélection.

Hormis les différences du « type » ou du « genre » du texte, on observe parfois des différences entre des lectures différentes d'un même texte : si, comme le soutient Eco, une œuvre comme *Ulysses* « peut être interprétée de différentes façons sans que son irrédicible singularité en soit altérée »⁸⁴, il semble que les différentes interprétations mettent en jeu des modes de sélection différentes. Todorov a d'ailleurs fameusement identifié un genre de texte – le « fantastique » – qui possède exactement cette propriété de permettre des interprétations différents (l'une « réaliste », l'autre « supranaturelle »)⁸⁵.

Enfin, pour prendre un exemple plus spéculatif, il se peut que les modes de sélection impliqués dans les lectures de textes de fiction aient évolué au cours de l'histoire. La question est de savoir si les modifications des lectures qui ont vraisemblablement accompagnés les évolutions indubitables qu'a connues le roman peuvent être toutes expliquées en termes des modifications (indubitables) des croyances, ou s'il y avait également des modifications des modes de sélection. Quand Barthes distingue la lecture qui convient au « texte moderne » de la lecture des « textes classiques »⁸⁶, s'agit-il d'une différence dans les croyances des lecteurs ou des modes de sélection employés ? Quand Pavel décrit le geste d'immerger une perfection naguère idéalisée dans un contexte réaliste qui constitua un des moments fondateurs de « la grande réforme de l'idéalisme romanesque »⁸⁷, le changement dans la lecture qui y va apparemment de paire, s'explique-t-il entièrement en termes des changements dans les croyances de la société à cette époque, ou plutôt par des changements dans les modes de sélection employés ? En vérité, il s'agit probablement d'un mélange des deux, mais cela suffirait pour mettre en évidence le phénomène difficile pour les théories de *faire comme si*, à savoir la possibilité de modifications du mode de sélection⁸⁸.

« modes ») de texte, concède un rapport entre ces derniers, lesquels sont « réalisés suivant les données des codes historiques » et l'expérience du lecteur (voir notamment pp358-362). La question plénière du rapport entre les « codes » et ce que l'on appelle ici « mode de sélection » est tangentielle aux propos de cet essai.

⁸⁴Eco (1965, p17, et p23).

⁸⁵Todorov (1970).

⁸⁶Barthes (1973, p18-21).

⁸⁷Pavel (2003, 2^e Partie, notamment Ch. III).

⁸⁸Walton (*op cit.*, §4.6) discute et rejette le versant sémantique de ce problème, qui met en question la stabilité de la notion de fiction et donc sa convenance comme « soubassement approprié à n'importe quelle théorie » (p185). Or, puisqu'il rejette ce problème en recourant à sa notion d'imagination (p185), il ne réussit pas à soulager les inquiétudes à l'égard de la stabilité de cette notion psychologique elle-même.

En résumé, on a mis en évidence trois difficultés qui touchent à la théorie des États Mentaux spécifiques à la fiction de Walton et Currie. Premièrement, il s'avère que la notion de *faire comme si* ou d'imagination qu'ils emploient *n'est pas suffisamment* sensible aux complexités des rapports avec les États Mentaux usuels à des moments particuliers. Une deuxième insuffisance concerne le rapport entre ces États Mentaux et les expériences phénoménales du sujet, qui n'est peut compris, surtout à ces mêmes moments. Ces rapports et ces moments ont une importante capitale pour le déroulement de la lecture ou de la considération de la fiction, à telle enseigne que ce manque de la part des théories de Currie et de Walton constitue un défaut grave dans leur compréhension du rapport entre réalité et fiction (dans la psychologie du lecteur). Troisièmement, dans la mesure où les modalités du rapport entre l'imagination ou le *faire comme si* et la croyance risquent de varier relativement aux textes et aux lectures, la notion de *faire comme si* risque d'être *trop* sensible aux rapports avec d'autres États Mentaux pour lui permettre de constituer une notion stable, appropriée à fonder une théorie.

La notion de *faire comme si* ou d'imagination proposée par Walton et Currie est comme la vitesse moyenne en physique : elle n'est pas assez fine ou sensible pour rendre compte des particularités pour lesquelles il faut la vitesse instantanée ; elle n'est pas assez générale pour être stable. Il reste seulement sa valeur heuristique.

1.4.2 Difficultés des États Mentaux usuels

Où il s'avère que des difficultés analogues à celles qui ont été mises en évidence pour les États Mentaux spécifiques à la fiction touchent aussi aux États Mentaux usuels, tels que la croyance. À savoir, on ne comprend pas bien, d'abord, les rapports entre différents États Mentaux usuels à certains moments, ensuite, le rapport entre ces États Mentaux et l'expérience phénoménale à ces moments, et enfin, les apparentes variations des modalités des États Mentaux d'une situation à l'autre.

L'État Mental de *faire comme si* n'est pas capable, à tout moment, de rendre compte des rapports précis entre fiction et réalité (dans la psychologie du lecteur), ni de son expérience phénoménale. L'État Mental de *faire comme si* n'est pas une notion stable ; il pourrait adopter des modalités différentes selon la situation. Quelle que soit sa valeur heuristique et son attraction intuitive, l'État Mental de *faire comme si* n'est donc pas à la hauteur de la question détaillée du

rapport entre fiction et réalité (dans la psychologie du lecteur) posée dans §1.1. Il s'avère cependant que de pareilles difficultés touchent *également* aux États Mentaux *usuels* (telle la croyance). A savoir :

1. il y a des moments particuliers où les rapports entre les États Mentaux *usuels* sont complexes et peu compris (tels un moment où l'un des États Mentaux est intuitivement *hors sujet*, un moment où l'un des États Mentaux entre en jeu, ou un moment d'hésitation et de délibération avant une décision) ;
2. les rapports entre l'expérience phénoménale et les États Mentaux (usuels) à ces moments sont également peu clairs ;
3. à l'encontre de la position d'un État Mental un (la croyance), il semble y avoir des variations possibles dans les modalités de ces États Mentaux entre des situations différentes.

Il importe d'insister sur cette communauté de difficultés, car elle indique que le défaut, pour ainsi dire, ne se situe pas complètement au niveau des États Mentaux spécifiques proposés (*faire comme si*, imagination), mais tient aux États Mentaux en général. Pour penser à fond la fiction, il ne convient pas de travailler en termes d'États Mentaux. Il s'agira donc, dans les chapitres ultérieurs, de développer des outils conceptuels plus fructueux, lesquels permettront une compréhension approfondie des complexités ayant trait à la lecture de la fiction.

Dans le paragraphe actuel, il s'agit de parcourir les trois difficultés qui se posent pour les États Mentaux usuels. Dans la mesure où l'on répète les arguments développés dans le paragraphe précédent, et parfois bien connus à la philosophie contemporaine, on se permet de passer assez vite sur les détails. Le seul point sur lesquels on s'attarde est l'application de ces arguments aux conceptions *sophistiquées* d'États Mentaux⁸⁹, laquelle n'a pas été discutée dans le paragraphe précédent puisque les États Mentaux spécifiques à la fiction ne sont pas sophistiqués (§1.3.2). Les conceptions sophistiquées d'États Mentaux, telles celle proposée par la théorie de la décision, sont souvent louées comme des « précisions rigoureuses » des conceptions ordinaires employées quotidiennement ; or, il découle du fait qu'elles tombent sous les mêmes difficultés qui touchent aux conceptions ordinaires que leur genre de « précision » n'est pas fécond pour penser les questions les plus épineuses de la fiction.

La première difficulté a trait au rapport entre différents États Mentaux (usuels), lequel est difficile et peu compris à certains moments. Aux trois types de rapports problématiques entre États Mentaux spécifiques à la fiction et

⁸⁹Voir §1.3.2 sur la distinction États Mentaux ordinaires et sophistiqués.

États Mentaux usuels⁹⁰ qui ont été présentés comme exemples dans le paragraphe précédent, semblent correspondre trois types de rapports également problématiques entre États Mentaux usuels. Il y a des moments où un État Mental est *hors sujet* et un autre État Mental est *en jeu* (cf. l'exemple des États Mentaux à l'égard de l'existence de Holmes) ; il y a des moments où des États Mentaux entrent en jeu (cf. la mise en jeu des croyances authentiques dans la lecture) ; il y a des moments d'hésitation avant une décision (cf. la réponse du lecteur à la question qui rompt sa lecture). Dans tous ces cas, les théories des États Mentaux ne rendent apparemment pas compte du rapport « effectif » entre les États Mentaux à ces moments. Ce qui apparaît clairement dans une brève considération de quelques exemples de tels moments.

Le premier type de rapport difficile est mis en évidence par la question suivante : le fait qu'un camarade de classe de l'école primaire, que vous avez oublié depuis des années, a des yeux bleus, est-ce toujours une de vos croyances ? Et quel rapport avait-elle avec vos croyances relatives à l'argumentation du paragraphe précédent ? Auparavant, cette croyance était *hors sujet*, si bien que le rapport à des croyances qui étaient en jeu (les propos du passage précédent) est complexe : on est tenté de poser des *priorités* différentes sur ces croyances, mais les théories des États Mentaux ne comprennent généralement pas les moyens de se représenter de telles « priorités ». Une deuxième difficulté s'attache aux rapports entre croyances aux moments où une d'entre elles entrent en jeu : dans quel sens et comment la croyance qu'un parapluie protège de la pluie, laquelle est normalement de *moindre priorité* ou *non active* entre-t-elle en jeu ou s'active-t-elle ou gagne-t-elle de priorité dans une situation où l'on croit qu'il pleut et où l'on veut sortir ? Généralement, les théories des États Mentaux admettent que cette croyance a été toujours « présente » dans le « corpus doxastique » de l'agent, mais elles n'expliquent généralement pas comment elle sort (ou est sortie) du « stockage » pour être mise à l'œuvre. Pour illustrer le troisième rapport problématique, il suffit de prendre les moments d'hésitation avant une décision (ou avant la réponse à une question), où l'on « pèse » les différentes croyances (et désirs) les unes contre les autres : l'attribution simple des États Mentaux qui rend compte de sa décision finale ne semble pas bien représenter la complexité des rapports entre États Mentaux impliqués dans cette hésitation ou délibération⁹¹.

⁹⁰Conforme aux affirmations antérieures, on ne considère que le cas des États Mentaux cognitifs.

⁹¹Plus précisément, il y a deux possibilités (supposant qu'il n'y a pas d'informations entrant en jeu entre l'hésitation et la décision) : tantôt on attribue les États Mentaux (croyances et désirs) à ce moment sur une base qui inclut la décision finale, de telle sorte que l'hésitation du

Il est nécessaire souligner la *généralité* de ce genre de difficulté : elle touche à toute conception d'État Mental, y compris les États Mentaux sophistiqués, tels ceux posés par la théorie de la décision (§1.3.2). Certains objecteraient que l'attribution des États Mentaux proposée par la théorie de la décision peut rendre compte du manque d'activité de la croyance relative au camarade de classe pendant la plupart de votre lecture de cet essai (par exemple), dans la mesure où vos comportements effectifs sont identiques à ceux que vous effectueriez si vous n'aviez pas cette croyance, selon les règles de la théorie de la décision et les États Mentaux que l'on vous a attribués⁹². Or, même si c'était le cas⁹³, la sensibilité de l'équilibre entre comportements contrefactuels et effectifs, la dépendance de toutes les croyances du sujet⁹⁴, pour ne pas mentionner l'appui sur un contrefactuel d'apparence difficile, empêcheraient d'appeler cette attribution une *compréhension* du rapport entre votre croyance *hors sujet* et vos croyances « au fleur de l'esprit ». C'est une compréhension seulement dans le sens où le fait que l'on puisse rendre compte du découlement de la Seine en termes de la mécanique quantique soit une compréhension. Cette complexité de la « compréhension » promise par la conception sophistiquée des États Mentaux ne pourrait que signaler un *défaut de fécondité* de cette approche du mental. En conséquence, une approche féconde de la fiction ne saurait pas se façonner d'après la théorie de la décision.

sujet n'est pas comprise (c'est le cas décrit dans le texte), tantôt on les attribue de telle manière qu'elles sont dans un équilibre fin (sur la base des comportements précédents seulement, par exemple), si bien que l'on comprend l'hésitation du sujet mais non pas la décision finale. Dans les deux cas, l'attribution manquera toujours un aspect de la complexité. Pour un autre exemple d'une situation qui ne permet apparemment pas d'attribution de croyances et désirs, voir l'exemple de l'erreur discuté par Dennett (1987, Ch. 4).

⁹²Cette approche de ce phénomène suppose que le problème consiste à rendre compte d'une croyance qui est *indépendante* des comportements actuels ; or, il n'est pas évident que l'*indépendance* (manque d'effets réciproques) et le *hors sujet* (ne pas être *en scène*) sont les mêmes choses. Au contraire, comme on montrera dans §4.2.4, le *hors sujet* et l'*indépendance* sont deux notions bien distinctes.

⁹³Il y a des doutes légitimes à l'égard de la capacité de la théorie de la décision de supporter une telle attribution. En bref, pour faire cadrer (tous) les comportements du sujet avec les règles (strictes) de la théorie de la décision, il faut se contenter d'une attribution « approximative » (voir par exemple Lewis (1974, p118), Davidson (1984, Ch. 9), mais également Quine (1960, Ch. 2)), de sorte que l'on risque de perdre les complexités les plus subtiles, telles les hésitations sur des décisions. Voir aussi §2.5.3.

⁹⁴Puisque l'action prise est, selon la théorie de la décision, celle qui maximise la somme sur les mondes possibles du produit du degré de croyance et de l'utilité, le « calcul » de l'action appropriée dépend des degrés de croyance relatif à chaque monde, c'est-à-dire, pour le dire grossièrement en français ordinaire, du degré de croyance relative à chaque énoncé du langage. Plus généralement, rappelons de §1.3.2 que les États Mentaux sophistiqués sont saturés et ont une base d'attribution globale.

La deuxième difficulté rencontrée par les États Mentaux usuels concerne le décalage entre l'expérience phénoménale et les États Mentaux attribués à certains moments ; de même que dans le cas d'États Mentaux spécifiques à la fiction considéré précédemment, elle ne consiste qu'en une réflexion sur le plan phénoménologique des difficultés touchant aux rapports entre États Mentaux. La croyance relative aux yeux du camarade de classe était *phénoménologiquement* absente avant qu'elle ait été mentionnée, comme elle était intuitivement *non active* et *hors sujet*. De même, *approximativement*, pour la croyance relative au parapluie. Dans le cas d'une décision ou d'une hésitation, la complexité des rapports semble bien *présente phénoménologiquement* dans l'expérience du sujet. Dans tous ces cas, les États Mentaux attribués ne correspondent pas à ces aspects de la phénoménologie. Les remarques faites dans le paragraphe précédent à l'égard de cette difficulté « phénoménologique » et de son rapport à la difficulté concernant les rapports entre États Mentaux, valent également dans le cas des États Mentaux usuels ; il ne convient pas de s'y attarder.

La troisième et dernière difficulté a trait aux éventuelles mutations des modalités des États Mentaux entre les circonstances, lesquelles ressemblent aux variations reconnues aux modalités du *faire comme si*. Il s'agit non seulement du fait que les dites croyances puissent s'attribuer sur la base des suites de comportements différentes selon la société et l'époque, mais, de plus, d'éventuelles sensibilités des attributions des croyances au contexte (un comportement qui soutiendrait l'attribution d'une croyance dans une situation ne la soutiendrait pas dans une autre), lesquelles sont reconnues depuis Wittgenstein⁹⁵. Dans la mesure où le « sens » ou (pour reprendre la terminologie wittgensteinienne) l'usage d'un État Mental est relatif au bloc comportant les autres États Mentaux avec lesquels il est attribué et les comportements qui soutiennent l'attribution, le « sens » de l'État Mental varie d'une situation à l'autre. Les États Mentaux – qu'ils soient spécifiques à la fiction ou usuels – sont susceptibles à des modifications de sens (et d'usage) d'une situation à l'autre⁹⁶.

⁹⁵Wittgenstein (1953, p161) : « Décrire mon état d'âme (de peur, par exemple [ou également, d'après Wittgenstein, de croyance]), je le fais dans un contexte particulier. ...Est-ce, alors, tellement surprenant que je me serve de la même expression dans différents jeux [de langage] ? » (pour Wittgenstein, les modalités d'attribution des États Mentaux sont spécifiques aux jeux de langage). Voir également Dretske (1970) ou Nozick (1981) pour des considérations similaires, relatives à la connaissance.

⁹⁶Seulement les États Mentaux *ordinaires* admettent la possibilité de ces modifications de sens, en tant que différences entre des blocs d'attributions différents (§1.3.2), alors que, dans l'attribution des États Mentaux sophistiqués, toutes les éventuelles modifications sont émoussées dans le processus d'ajustement et d'approximation par lequel les règles de la théorie de la décision sont « appliquées » aux comportements du sujet (voir la note 93).

Reste que les États Mentaux traditionnels – de croyance, de désir, des émotions – sont très utiles pour certains propos. Peut-être qu’il en soit de même pour la notion de *faire comme si* ou d’imagination, malgré la faible compréhension du rapport avec les autres États Mentaux, et leur air *ad hoc*. Pourtant, pour ce qui concerne les questions détaillées et difficiles posées ici, la notion de *faire comme si* (ou d’imagination waltonienne), ainsi que des États Mentaux usuels, ne suffisent pas. Ce qui appelle un recours aux mécanismes supposément plus *fins*, dont notamment les diverses propositions de « dispositifs psychologiques ». Là encore, on verra que les mécanismes proposés pour rendre compte de la fiction sont insuffisantes.

1.5 Approches « causales » : mécanismes spécifiques à la fiction

Les questions détaillées à l’égard de la fiction posées dans §1.1 s’avèrent trop subtiles pour les États Mentaux grossiers de *faire comme si* ou d’imagination proposés par Walton et Currie ; par conséquent, il faut se tourner vers les enquêtes sur les aspects dits « causaux » (§1.3.1), qui ne se contentent pas avec les attributions des attitudes grossières par un observateur mais qui prétendent au contraire s’occuper des « mécanismes réels » sous-jacents au comportement humain. Comme souligné dans §1.3.1, il s’agira seulement de la *fécondité* des *mécanismes* par rapport à la question détaillée et « internaliste » posée relative à la fiction (§1.1.2), et nullement de leur *réalisme physique* ou de leur *opérabilité informatique*. L’intérêt des mécanismes discutés ici ne consiste *en rien d’autre* que leur prétendue *finesse de structure* et leur capacité de rendre compte des complexités de la fiction en termes de cette structure.

En pratique, tous les mécanismes ou systèmes psychologiques qui abordent d’une manière ou d’une autre la question de la fiction adoptent la stratégie que l’on a qualifiée de « mimétique » (§1.2.1). C’est-à-dire, ils commencent tous par poser une distinction claire et tranchante entre un côté fictionnel et un côté réel, pour ensuite restituer d’une manière ou d’une autre un degré de similitude entre les deux côtés séparés. On espère ainsi rendre compte des deux aspects de la relation entre fiction et réalité mis en relief dans §1.1.3. Toutefois, en fin de compte, ces mécanismes sont touchés par les mêmes faiblesses que l’on a repérées dans le cas des États Mentaux spécifiques de *faire comme si* ou d’« imagination ». D’abord, les rapports complexes entre les mécanismes fictionnels et les mécanismes « usuels » ou « réels » sont très mal compris, surtout à certains moments particuliers ; ensuite, le rapport entre les mécanismes proposés et les

expériences phénoménales du sujet est également difficile et peu compris, surtout à ces mêmes moments. Enfin, les mécanismes ont peine à rendre compte des modifications éventuelles dans les modalités du rapport entre fiction et réalité.

Ce sont des *faiblesses structurelles* des mécanismes proposés, qui mettent en question leur suffisance pour aborder les questions épineuses et difficiles du rapport entre fiction et réalité. Or, comme dans le cas des États Mentaux, les difficultés rencontrées touchent également aux théories des mécanismes « ordinaires » relatifs aux compétences humains d'ordre supérieur, telles le raisonnement. Il convient donc de chercher un autre *cadre* psychologique plus adéquate à ces questions ; d'ailleurs, une analyse des difficultés rencontrées par les théories proposant ces mécanismes pointe le chemin vers un tel cadre, chemin qui sera emprunté dans les chapitres suivants. Comme anticipé dans §1.2.3, ce cadre se prêtera à une approche non mimétique à la fiction (au sens de §1.2.1).

On présente d'abord quelques mécanismes suggérés qui ont trait à la fiction, et montre qu'ils adoptent tous la stratégie « mimétique » à son égard. Dans un deuxième temps, on considère les faiblesses communes à toutes ces propositions. Ensuite on remarque brièvement que des considérations similaires s'appliquent aux mécanismes proposés relatifs aux comportements « ordinaires » du sujet. Ces considérations permettent enfin une analyse générale de ces difficultés, laquelle ouvre la voie vers une approche mûre et sophistiquée de la question de la fiction.

1.5.1 Quelques mécanismes psychologiques

Où l'on présente quelques mécanismes psychologiques relatifs à la fiction proposés dans la littérature. De manière générale, ils s'opposent sur la façon d'établir la distinction, dans le mental, entre les aspects fictionnels et réels : certains conçoivent une distinction selon le statut des entités mentales, d'autres selon les processus impliqués, et d'autres encore admettent une distinction aussi bien selon le statut que selon le processus. Ces théories ont cependant en commun le fait d'employer la stratégie « mimétique » (§1.2.1).

Il y a dans psychologie cognitive un vif débat portant sur les mécanismes mentaux relatifs à la *feinte* (*pretense*), au sens *ludique* de ce terme. L'intérêt de ce débat pour les propos de l'étude actuelle réside dans le rapport entre la *feinte* (ludique) et la fiction, rapport déjà évident dans l'approche de la fiction proposée par Walton et Currie, mais également dans l'approche schaefferienne⁹⁷.

⁹⁷Schaeffer conçoit le cadre pragmatique relatif à la fiction comme celui d'une « feintise

Ce rapport est souvent mobilisé pour mettre en rapport le débat sur la *feinte* dans la théorie d'esprit et la question psychologique relative à la fiction⁹⁸ ; on l'admet et on s'en sert pour les propos de la discussion suivante.

Le débat autour de la feinte (qui sera exposé ici en tant qu'il s'applique à la fiction) oppose typiquement deux conceptions : la théorie dite « théorique » (*theory theory*) et la théorie dite « simulationniste » (*simulation theory*). La version première et typique de la première conception a été formulée dans Leslie (1987), mais des approches apparentées incluent Plagnol (1993, 2004) et (dans une certaine mesure) Nichols & Stich (2003). La deuxième conception compte parmi les articles pionniers Heal (1986), Gordon (1986) et Goldman (1989), auxquels se joignent les travaux de Currie (1995) et de Harris (1992)⁹⁹. Ces théories *s'accordent* à considérer le mental comme un ensemble de *mécanismes* (au sens le plus large) qui *opèrent* sur des *éléments*, que l'on conçoit souvent comme des « représentations » et que l'on appelle ici des « entités mentales »¹⁰⁰. (On adopte ici la convention de référer aux entités mentales en employant l'expression linguistique en majuscules qui réfère à ce que les entités « représentent » – HOLMES pour l'entité qui « représente » Holmes, par exemple.) Les théories *s'opposent* sur la distinction entre fiction et réalité dans le mental :

- les théories « théoriques » établissent la distinction selon le *statut* des entités mentales : certaines entités mentales ont un *statut fictionnel* (HOLMES EXISTE) et d'autres ont un *statut ordinaire* (GLADSTONE EXISTE) ;
- les théories « simulationnistes » établissent la distinction selon le *processus* mental approprié : certains emplois de certains mécanismes mentaux, c'est-à-dire l'effectuation de certains processus mentaux, sont *fictionnels*, alors que d'autres sont *ordinaires* (le processus mental relatif à l'inférence que Gladstone exista à partir des propos d'une œuvre historique tenu pour vrai est un processus *ordinaire*, alors que le processus qui *simule* ce premier, en inférant que Holmes existe à partir des propos des *Aventures de Sherlock*

ludique partagée » (1999, Ch. III, §2). Selon Walton et Currie, l'activité propre à la fiction est l'imagination dans le jeu de *faire comme si* ; autrement dit, la *feinte* ou le *faire semblant* dans ce jeu (§1.2.1).

⁹⁸Voir particulièrement Currie (1995).

⁹⁹Le recueil de Davies and Stone (1995) présente un aperçu de plusieurs positions simulationnistes.

¹⁰⁰Certains (Leslie (1987), Nichols & Stich (2003), Currie (1995), par exemple) conçoivent ces entités sous le mode propositionnelles (structurellement semblables aux énoncés du langage) ; d'autres (Plagnol (1993, 2004), dans la tradition d'Anderson (1983) et Kintsch (1988)) permettent que des entités peuvent être également conceptuelles (structurellement semblables, *grasso modo*, aux mots du langage). On aura ici recours aux deux conceptions des entités selon l'auteur discuté.

Holmes, est un processus *fictionnel*).

Pour illustrer ces deux distinctions, on présente brièvement quelques théories soutenant l'une ou l'autre, aussi bien que quelques théories « hybrides » qui distinguent entre fiction et réalité (dans le mental) selon le *statut* aussi bien que le *processus*.

Le premier exemple d'une théorie qui fait la distinction entre fiction et réalité selon le *statut* des entités mentales est celle de Leslie (1987), selon laquelle les entités mentales « fictionnelles » se distinguent des entités mentales « ordinaires » par leur « structure interne », de la même manière que, selon une hypothèse répandue, les énoncés « fictionnels » se distinguent des énoncés ordinaires par le fait d'être préfixés par un opérateur « dans la fiction, ... » (§1.2.1). On obtient l'entité mentale correspondant au fait fictionnel que Holmes existe en deux étapes : pour l'entité mentale **HOLMES EXISTE**, on permet d'abord une entité mentale qui la « représente » ('**HOLMES EXISTE**'), à la façon dont la citation d'un énoncé ('Holmes existe') « réfère » à l'énoncé lui-même (Holmes existe) ; ensuite, on se joint à cette entité mentale l'opérateur mental **DANS LA FICTION**, pour aboutir à l'entité mentale **DANS LA FICTION 'HOLMES EXISTE'**, qui, selon Leslie, est l'entité mentale qui « représente » le fait que dans la fiction, Holmes existe¹⁰¹. Il y a évidemment une différence de structure et donc de *statut* entre **GLADSTONE EXISTE** et **DANS LA FICTION 'HOLMES EXISTE'** : la première entité mentale est « ordinaire », la deuxième est « fictionnelle ».

Plagnol (1993, 2004) et Nichols & Stich (2003) tirent également des distinctions entre les entités mentales selon leurs *statuts*, mais à la différence de Leslie, il s'agit d'une distinction selon leur *position* dans la *structure* qui relient les entités mentales plutôt que d'une distinction selon la « structure interne » de l'entité mentale elle-même. Selon Plagnol, les entités mentales sont organisées dans une structure à la façon des nœuds dans un réseau, et cette structure contient une « entité mentale » correspondant à un texte (de fiction) donné, laquelle est reliée aux entités qui relèvent de la lecture de ce texte : par exemple, l'entité men-

¹⁰¹Voir Leslie (1987, p419 *sq.*), pour la présentation de sa théorie psychologique. Les considérations sémantiques qui servent à motiver sa théorie psychologique (p416-419) emploient la stratégie « mimétique », dans la mesure où elles mettent en jeu un opérateur « dans la fiction, ... » (ou, dans le cas de la feinte, « X feint que ... »). La méta-représentation de l'entité mentale (conçue sur le modèle de la citation linguistique) est appelée, selon Leslie, par l'*opacité* (Quine, 1960, §30*sq.*) de l'opérateur « il feint que », ou également « dans la fiction ... », c'est-à-dire le fait de ne pas permettre la substitution des termes co-référentiels *salve veritate* (comparer « dans la fiction, Holmes habite Londres » et « dans la fiction, Holmes habite la cité qui accueillit les Jeux Olympique de 2012 »).

tales AVENTURES DE SHERLOCK HOLMES et reliée à l'entité mentale HOLMES¹⁰². Pour Nichols et Stich, qui pensent les entités mentales comme rangées dans des « boîtes » selon qu'elles sont les croyances ou les désirs (par exemple), il y a, à côté de la « boîte de croyances », une « boîte de mondes possibles » qui « contient » les entités mentales correspondant à des énoncés « crus relatifs à la fiction », telles l'entité mentale HOLMES EST UN DETECTIVE¹⁰³. Dans les deux cas, il y a une *compartimentation* de l'ensemble des entités mentales ayant trait à un texte de fiction, qui les « sépare » et les « distingue » des entités mentales « ordinaires ».

À l'opposé, pour venir aux distinctions établies selon le *processus*, les théories « simulationnistes » développées par Gordon (1986), Goldman (1989), mais également par Currie (1995)¹⁰⁴, soutiennent que les *processus* mentaux relatifs à l'activité de feinte, ou à l'imagination dans la fiction, se ressemblent à des *processus* mentaux ordinaires, à ceci près que certaines connexions ordinaires à d'autres mécanismes ou processus mentaux manquent. Pour l'exprimer par une devise brève, qui a du moins le mérite de souligner l'aspect *mimétique* des théories de la simulation, la lecture de la fiction est une « simulation mentale » d'une lecture ordinaire, où une « simulation mentale » est « un processus mental qui est, ou est censé être, *isomorphe* dans certains aspects pertinents au processus cible qu'elle est censée imiter »¹⁰⁵. Par exemple, pour Gordon et Goldman, il s'agit de faire opérer certains mécanismes *ordinaires* d'inférence, mais avec deux différences : premièrement, on leur fournit des opérandes éventuellement fictionnels ou contrefactuels, en l'occurrence les propos du texte de fiction ; deuxièmement, certains résultats de ces mécanismes sont « déconnectés » des processus conséquents ordinaires, tels les comportements « normalement » impliqués si le lecteur croyait que le texte était véridique¹⁰⁶. La fiction se distingue de la réalité (dans le mental) par cette différence dans les processus ou les opérations des mécanismes : l'une est « authentique », l'autre la simule.

¹⁰²Pour des discussions de la gestion des entités mentales relatives à la fiction ou à une histoire particulière, voir Plagnol (1993, p43 *sq*) et (2004, Ch II, §2), ou Potts & Peterson (1985).

¹⁰³Voir par exemple Nichols & Stich (2003, §2.4.1), où ils traitent la question de la feinte.

¹⁰⁴Currie (1995) fait le rapport explicite entre la simulation et l'imagination relative à la fiction (« l'imagination est le simulateur », p158), mais il n'y a pas le seul à le faire : Feagin, dans un ouvrage centré sur la question de la lecture de la fiction, s'appuie fortement sur les théories « simulationnistes » (Feagin, 1996, Ch. 4).

¹⁰⁵Goldman (1989, p108).

¹⁰⁶On dit souvent que les mécanismes s'opèrent de manière « autonome » (*off-line*) ; voir Gordon (1986) et Goldman (1989). Currie, *op cit.*, semble s'opposer à cette position, en posant un « simulateur » dans la machinerie mentale qui opère la simulation.

Enfin, certaines théories posent (explicitement) une distinction entre fiction et réalité aussi bien selon le *statut* des entités mentales relevant de l'une et de l'autre que selon les *processus* relatifs à l'une et à l'autre. La théorie développée par Nichols & Stich, selon laquelle *d'une part* le processus de la lecture de la fiction est une *simulation* du processus ordinaire de lecture, et *d'autre part* les états cognitifs produits par la simulation sont mis dans la « boîte de mondes possibles » plutôt que la « boîte de croyances », en est un exemple¹⁰⁷. Un autre exemple important, formulé spécifiquement pour la fiction, est la théorie récemment proposée par Schaeffer. D'un côté, les entités mentales (« représentations ») relatives à la fiction appartiennent à « un espace de jeu à l'intérieur duquel ...[elles] ne sont traitées de la même manière que le seraient les représentations « réelles » mimées par le dispositif fictionnel »¹⁰⁸, mais elles ressemblent aux entités mentales « usuelles », dans la mesure où elles sont exactement des mêmes classes et des mêmes modalités¹⁰⁹. Mais de l'autre côté, Schaeffer tend à concevoir la croyance ordinaire comme l'achèvement d'une longue démarche « pré-attentionnelle », dont toutes les étapes sauf les dernières sont parcourues dans le cas de la fiction¹¹⁰, ce qui lui mène à concevoir la différence entre les entités mentales « fictionnelles » et « ordinaires » comme conséquence d'une différence dans les processus qui leur aboutissent (dans le cas ordinaire, le processus termine en posant une croyance, dans le cas de la fiction, il s'achève avant cette étape). En ce sens, Schaeffer propose une distinction entre fiction et réel dans le mental aussi bien selon le *statut* de l'entité mentale que le *processus* mental impliqué.

Il n'est pas difficile à voir que toutes ces théories adoptent la stratégie « *mimétique* » (§1.2.1), selon laquelle on pose d'abord une distinction entre un « côté » de fiction et un « côté » de réalité, pour ensuite récupérer des similitudes entre les deux « côtés » distingués. Les entités mentales « fictionnelles » de Leslie diffèrent des entités mentales par le fait de figurer l'opérateur **DANS LA FICTION**, mais ils y ressemblent dans la mesure où '**HOLMES EXISTE**' « représente » **HOLMES EXISTE** qui est du même genre qu'une entité mentale ordinaire. Les entités mentales « fictionnelles » de Plagnol ou de Nichols & Stich

¹⁰⁷Voir notamment Nichols & Stich (2003, Ch. 3) où ils traitent la question de l'attribution des attitudes à l'autrui.

¹⁰⁸Schaeffer (1999, p 162, mais voir aussi p 153-156).

¹⁰⁹Schaeffer, *op cit.*, p153 *sq.* : « Il faut donc abandonner l'idée selon laquelle il existerait deux modalités de représentation, l'une qui serait fictionnelle et l'autre qui serait référentielle ».

¹¹⁰Schaeffer, *op cit.*, p 189 : « La situation d'immersion fictionnelle se caractérise donc par l'existence conjointe de leurres mimétiques pré-attentionnels et une neutralisation concomitante de ces leurres par un blocage de leurs effets au niveau de l'attention consciente. »

se distinguent des entités mentales « ordinaires » par leur « position mentale » (relié à l'entité mentale **LES AVENTURES DE SHERLOCK HOLMES**, dans la « boîte de mondes possibles »), mais, à part ça, elles s'y ressemblent parfaitement. Les processus impliqués dans la lecture de la fiction sont bien, selon Currie et les simulationnistes, des simulations plutôt que des processus « ordinaires », mais comme telles elles sont « isomorphes » aux processus « ordinaires ». Et selon Schaeffer, non seulement les entités mentales « fictionnelles » et « ordinaires » se distinguent par leur appartenance à un certain « espace de jeu » et se ressemblent autrement, les processus qui les ont introduites se distinguent par leur dernière étape (position de croyance ou non) et se ressemblent autrement. Ayant adopté cette stratégie « mimétique », il incombe à ces théories, si elles veulent répondre à la question détaillée du rapport entre la fiction et la réalité (§1.1.3), de préciser exactement le rapport entre les entités mentales ou processus « fictionnels » et les entités ou processus « ordinaires ». Il s'avère qu'elles n'ont pas les ressources pour ce faire, et cela pour les mêmes raisons.

1.5.2 Les limites des théories proposant des mécanismes spécifiques à la fiction

Où l'on voit que les trois difficultés rencontrées par les États Mentaux spécifiques à la fiction, ayant trait aux rapports entre fiction et réalité à des moments particuliers, aux rapports entre les aspects psychologiques proposés et l'expériences phénoménales du lecteur, et aux modifications des modalités de la lecture, se posent également pour les mécanismes spécifiques à la fiction. En outre, la manière dont les théories des mécanismes spécifiques à la fiction tentent de rendre compte de ces aspects difficiles, à savoir en les renvoyant à d'autres mécanismes, n'est guère suffisante. Car, d'une part, le recours à d'autres mécanismes met en question le besoin et l'importance des mécanismes spécifiques à la fiction proposés et, d'autre part, il est douteux que les mécanismes auxquels on fait appel offrent quelque chose de l'ordre d'une compréhension de ces phénomènes complexes et difficiles.

La théorie des *États Mentaux* spécifiques à la fiction (théorie de « *faire comme si* » de Currie, théorie de l'« *imagination* » de Walton) comporte trois lacunes importantes. Premièrement, on comprend mal les rapports entre les États Mentaux spécifiques du lecteur et ses États Mentaux usuels à certains moments importants. Deuxièmement, on comprend mal le rapport entre ces États

Mentaux et l'expérience phénoménologique à ces moments. Troisièmement, les éventuels *changements* dans les modalités du *faire comme si* ou de l'imagination sont à peine tolérés, et encore moins compris (§1.4.1).

Du côté des *mécanismes* « spécifiques » proposés comme sous-jacents à la lecture et à la considération de la fiction (différence de statut des entités mentales, simulation de processus « ordinaire »), ces lacunes se répètent. À savoir :

1. il y a des moments importants où des mécanismes « spécifiques » à la fiction entrent en jeu (moment d'immersion fictionnelle, d'entrée ou de sortie de cette immersion, de mise en jeu d'un fait réel dans la lecture) mais où leur relation entre les mécanismes avec les mécanismes « ordinaires » n'est pas comprise *par les théories qui les posent*, sans recourir à d'autres mécanismes mentaux.
2. la relation entre les mécanismes spécifiques à la fiction et la phénoménologie de l'expérience du lecteur est également mal comprise, à ces mêmes moments.
3. on comprend à peine comment le statisme induit par les structures fixes des mécanismes spécifiques à la fiction permet de rendre compte des possibles évolutions de la lecture (entre des textes, des lectures et éventuellement des époques différents).

Puisque ces difficultés sont à maints égards très semblables à celles qui ont été soulevées et discutées dans §1.4.1, on passera relativement vite sur les détails des exemples, en référant le lecteur à §1.4.1, pour se concentrer sur les aspects qui sont particuliers aux cas des mécanismes.

Rapports entre mécanismes « spécifiques » et « ordinaires » et le recours à un mécanisme « central »

Quant à la première difficulté, on a identifié dans §1.4.1 trois moments où certaines relations entre des États Mentaux spécifiques à la fiction et des États Mentaux ordinaires sont peu compris : d'abord, la relation entre la croyance à l'inexistence de Holmes et le *faire comme s'il existait*, à un moment d'immersion dans la lecture ; ensuite, la relation entre ces deux États Mentaux à moment où la lecture est rompue pour poser la question de l'existence de Holmes ; enfin, la relation entre les *faire comme si* et les croyances ordinaires qui entrent en jeu dans la lecture (les être humains ne survivent pas à des chutes d'escarpements élevés) aux moments de leurs entrées en jeu. Ces relations posent également des problèmes pour les mécanismes « spécifiques », qu'ils se distinguent des mécanismes « ordinaires » selon le statut ou selon le processus.

D'un côté, si l'on distingue la fiction et la réalité selon le *statut* des entités mentales, alors la relation entre les attitudes à l'égard de l'existence et de l'inexistence de Holmes, correspond, sur le plan des mécanismes, à une relation entre l'entité mentale « ordinaire » correspondant à la croyance à l'inexistence de Holmes et l'entité mentale « fictionnelle » correspondant au *faire comme s'il existait*. Or, la relation entre ces attitudes manifeste, à certains moments, une subtilité et une complexité dont il n'y a aucun équivalent sur le plan des entités mentales. Pour prendre l'exemple du moment de l'immersion dans la lecture, aucun aspect de la relation entre des entités mentales « fictionnelles » et « ordinaires », en tant qu'elle est présentée par les théories qui proposent cette distinction, ne rend compte de l'apparente *priorité* de l'existence (fictionnelle) de Holmes sur son inexistence (réelle) qui s'établit à ce moment (§1.4.1). Pour employer l'exemple de la question de l'existence de Holmes posée pendant la lecture, la position de deux entités mentales – l'entité **HOLMES EXISTE** avec un statut « fictionnel » et l'entité **HOLMES N'EXISTE PAS** avec un statut « réel » – ne suffit pas pour comprendre la complexité du rapport entre ces deux attitudes lors de l'hésitation et de la délibération du lecteur avant sa réponse (§1.4.1). Pour se tourner enfin vers la question de l'entrée en jeu des croyances réelles dans la lecture, là encore, rien n'est offert qui permet de comprendre quels nouveaux rapports sont instaurés (avec les entités mentales « fictionnelles ») lorsque l'entité mentale correspondant à la croyance « entre en jeu ».

De l'autre côté, si l'on distingue fiction et réalité selon les *processus* mentaux, c'est la relation entre les processus se déroulant « dans » la simulation (où Holmes existe) et les processus « ordinaires » se déroulant « dehors » (où Holmes n'existe pas) qui est en question. Or, encore une fois, rien du degré de complexité et de subtilité que l'on trouve dans les relations entre les attitudes n'est explicitement proposé entre les processus « fictionnels » et « ordinaires ». À un moment de la lecture, où il y a une sorte de priorité d'une attitude sur l'autre, une priorité « correspondant », du processus se déroulant « dans » simulation sur les mécanismes « ordinaires », n'est guère proposée ni point comprise dans les détails. De manière similaire, il n'y a aucune compréhension de la relation entre la simulation et d'autres mécanismes « ordinaires » qui peuvent rendre compte de l'hésitation et de la délibération lorsque le lecteur se voit poser la question de l'existence de Holmes pendant sa lecture. Enfin, les croyances ordinaires qui entrent en jeu dans la lecture sont, selon les simulationnistes, les opérands de la simulation ; or, ils ont à peine détaillé comment ces opérands sont sélectionnés pour la simulation, ainsi que quels opérands y sont fournis. Autrement dit, ils ont à peine compris les rapports entre les mécanismes « réels », qui ont affaire

à ces genres de croyances, et la simulation « fictionnelle » qui s'en sert.

En résumé, la simple distinction entre mécanismes « spécifiques » à la fiction et mécanismes « ordinaires » posée par ces théories n'offre pas les ressources pour comprendre le *rapport* des uns aux autres, notamment à des moments particuliers et d'ailleurs importants pour la compréhension des conduites du lecteur relatives à la fiction (§1.4.1). Or, à la différence du cas des États Mentaux, ces théories peuvent se permettre de renvoyer la question de ce rapport à d'autres mécanismes psychologiques « dans l'esprit », de telle sorte que, en théorie, il y a *quelque mécanisme* qui rend compte des relations complexes et difficiles entre entités mentales « fictionnelles » et « ordinaires » ou entre processus « simulés » et « authentiques », mais il n'appartient pas à ces théories de comprendre ce « quelque mécanisme ». Par exemple, pour Leslie, Nichols & Stich, ou peut-être Currie, qui adhèrent à une conception « modulariste » de l'esprit, où les mécanismes ayant trait à la fiction sont des « systèmes spécifiques » ou des dits « modules » avec des capacités limitées (tantôt pour opérer et gérer certains aspects des entités du statut fictionnel¹¹¹, tantôt pour effectuer les simulations¹¹²), il faudrait un système plus central¹¹³, appelé parfois « système directorial » ou « administrateur central »¹¹⁴, qui est responsable des tâches plus difficiles et complexes, telle la gestion du rapport entre les différents systèmes ou modules. Selon ces approches de la fiction, il est apparemment sous-entendu que *ce système directorial* s'occupe du rapport entre les entités ou les processus « fictionnels » et « réels ». D'un autre côté, Plagnol, adhérent à une conception plus « associationniste » de l'esprit comme un ensemble d'entités mentales reliées par des liens de force variable, soutient que les entités mentales sont dotées avec des niveaux d'*activation* selon le degré auquel les entités mentales sont « pertinentes » ou « en jeu »¹¹⁵. Dans cette optique, c'est la notion d'*activation* qui est censée permettre de rendre compte des rapports complexes entre les entités mentales « ordinaires » et celles qui sont « fictionnelles » (reliées à une entité FICTION) : la « valorisation » de l'existence (fictionnelle) de Holmes sur son inexistence (effective) aux moments de l'immersion pourrait être représentée par un niveau d'activation élevé de l'entité HOLMES EXISTE et un niveau minoré

¹¹¹Leslie, *op cit.*, Nichols & Stich, *op cit.*, §2.4

¹¹²Par exemple, le « simulateur » de Currie, *op cit.*

¹¹³Ou autrement *des* systèmes centraux, alors que, pour simplifier la discussion, qui vaut qu'il y ait un ou plusieurs systèmes centraux, on se limite au cas d'un système.

¹¹⁴Voir par exemple Fodor (1985, Ch. IV), mais également Johnson-Laird (1983, Ch. 16). Le terme « système directorial » (*Executive system*) vient de Schacter (1989), celui d'« administrateur central » (*Central Executive*), de Baddeley (1992).

¹¹⁵Plagnol, *op cit.*, p 27. Cette conception emprunte au travail de Anderson (1983, Ch. 3). En soulignant cet aspect, on laisse de côté une grande part de la théorie de Plagnol.

d'activation de l'entité HOLMES N'EXISTE PAS.

Ce recours à un autre mécanisme psychologique (au sens large) pour s'occuper des aspects complexes et difficiles relatifs à la fiction a deux conséquences importantes. Premièrement, il risque d'affaiblir la thèse selon laquelle les mécanismes proposés rendent compte, voire même sont nécessaires pour rendre compte, de la fiction en tant qu'elle apparaît dans le mental. Deuxièmement, les comportements du mécanisme auquel on a recours sont tellement complexes et difficiles qu'ils induisent des doutes à l'égard de la thèse selon laquelle ce mécanisme puisse fournir une *compréhension* de la fiction et de sa relation avec la réalité (en tant qu'aspects du mental).

Tout d'abord, il n'y a aucun doute que le recours à un autre mécanisme psychologique le charge des aspects les plus difficiles relatifs à la fiction : si la question de la relation entre entités mentales ou processus « fictionnels » et « ordinaires » est la question difficile et importante de la fiction, et si c'est à quelque autre mécanisme – qu'il soit « système directorial » ou partition de l'activation – d'établir et de gérer cette relation, c'est le système directorial ou la notion d'activation qui tient la clef de la question épineuse de la fiction, et nullement quelque distinction entre entités ou processus « fictionnels » et autres.

Ensuite, ce mécanisme « central » est lui-même complexe et difficile à comprendre. D'autant plus que, comme l'a été souligné dans §1.4.1, le vrai enjeu de la question du rapport entre les aspects « fictionnels » et « ordinaires » à des moments particuliers consiste en son importance pour la question de la *dynamique* de ce rapport, en telle sorte que la tâche qu'est censé accomplir ce mécanisme comprend non seulement l'établissement du rapport entre mécanismes « fictionnels » et « ordinaires » mais en outre la *gestion* de sa *dynamique*. Au stade actuel, les nombreux progrès dans la compréhension et la modélisation des mécanismes « centraux » ne semblent pas permettre une compréhension détaillée du rapport entre la fiction et réalité dans le mental. Et même si une compréhension détaillée et complète de ce genre de mécanisme était possible – même si l'étude des « systèmes directoriaux » ou de la dynamique de l'activation débouchait sur des principes précis¹¹⁶ –, l'application de ces principes à des individus particuliers et à des lectures particulières risque d'être tellement complexe,

¹¹⁶ Les modélisations de l'activation contiennent souvent des principes simples et précis qui gèrent les niveaux d'activation de nœuds (voir par exemple les équations d'Anderson, *op cit.*, p92-97); l'argument vise à suggérer que, même s'il était possible de spécifier tous les paramètres inconnus (B et p^* dans Anderson) et la structure du réseau (son r_{ij}), on ne saurait pas se dire *comprendre* les propriétés à grande échelle pertinentes pour les activités mentales « d'ordre supérieur », telles le raisonnement et la lecture de la fiction.

subtile et difficile, comportant une dépendance d'une telle gamme d'autres facteurs et comportements que l'on ne peut pas parler d'une *compréhension*. En un mot, même s'il y avait la *possibilité* de rendre compte du rapport psychologique entre fiction et réalité en ces termes, on aurait droit à douter la *fécondité* de ces approches et des « réponses » qu'elles prétendent pouvoir offrir¹¹⁷.

En somme, les mécanismes spécifiques proposés n'arrivent pas à rendre compte, de manière satisfaisante, des complexités du rapport entre fiction et réalité qui ont échappés aux États Mentaux spécifiques à la fiction. En particulier, le recours à un mécanisme « central » ne paraît guère suffisant : non seulement la nécessité d'un tel recours semble saper le soutien pour les mécanismes proposés comme spécifiques à la fiction et surtout le soutien pour la thèse selon laquelle ces mécanismes rendent compte de la fiction, mais il est douteux qu'un tel mécanisme puisse offrir quelque chose de l'ordre d'une *compréhension* du rapport entre fiction et réalité. Il s'avère que, quant aux autres difficultés mises en évidence dans §1.4.1, les mécanismes spécifiques à la fiction ont besoin encore du recours à un mécanisme central, de telle sorte que les considérations précédentes s'appliquent de nouveau.

Expériences phénoménales : rapports aux mécanismes spécifiques et au mécanisme central

Il n'est pas difficile de voir que la question de la deuxième faiblesse de la théorie des États Mentaux spécifiques se reproduit sur le plan des mécanismes. Aux moments pris pour exemples ci-dessus – le moment d'immersion fictionnelle, le moment où elle est rompue, le moment où un fait réel « entre en jeu » –, on a déjà vu dans §1.4.1 que la phénoménologie du lecteur correspond assez bien aux rapports entre les États Mentaux qui sont manqués par les théories du *faire comme si*. Mais on vient de voir qu'à ces moments, les rapports entre les mécanismes proposés comme « spécifiques » à la fiction et les mécanismes « ordinaires » ressemblent à ces rapports entre États Mentaux, et que les théories qui proposent ces mécanismes ont peine à en rendre compte ; en conséquence, la phénoménologie du lecteur leur échappe aussi.

L'identification de la phénoménologie comme un aspect, voire un aspect *important*, qui échappe à ces théories est par ailleurs une conséquence des théories relatives aux mécanismes « centraux » (système directorial, niveau d'activation) auxquelles ont recours les théories des mécanismes spécifiques à la fiction. Car ces théories admettent, sinon une « identification » entre la phénoménologie du

¹¹⁷Cet argument est analogue à celui développé dans §1.4.2 pour mettre en doute la capacité des États Mentaux sophistiqués de rendre compte de la fiction.

lecteur et les mécanismes « centraux », du moins un rapport étroit entre eux. Dans l'optique « modulariste », on admet que le système directorial a trait à quelque chose d'*attentionnel*, dans la mesure où les aspects sélectionnés par le système directorial sont *accessibles* aux processus tel le raisonnement dont le sujet est normalement conscient¹¹⁸. Même si l'on refuse d'identifier ce genre d'« attention » à l'expérience phénoménale, on admettrait quelque autre mécanisme (au sens large) qui ont trait à la phénoménologie, tel le système spécifique à la conscience phénoménologique (le *Conscious Awareness System*) proposé par Schacter, qui entre dans un rapport étroit avec le système directorial¹¹⁹. Également, dans l'optique « associationniste », on est habituellement d'accord sur le rapport entre l'activation et le degré d'*attention* – les aspects accordés un niveau d'activation élevé sont *pertinents* pour les processus mentaux¹²⁰ –, et même si l'on insiste, à la façon de Cowan, pour distinguer l'information mentale « activée » (munie d'un niveau d'activation non nulle) du *focus attentionnel* (qui contient les éléments dont le sujet est conscient), on admet qu'il y a un rapport important entre eux (en l'occurrence, que le dernier est un sous-ensemble du premier)¹²¹. En bref, on est arrivé au décalage entre l'expérience phénoménale et les mécanismes spécifiques à la fiction par une autre voie : le rapport difficile et complexe entre ces mécanismes est établi et géré, selon ces théories, par quelque mécanisme « central », qui est, lui, en rapport étroit avec la phénoménologie du lecteur.

Modifications de modalités de lecture et les mécanismes spécifiques à la fiction

La troisième difficulté qui pourrait se poser pour les théories des mécanismes spécifiques à la fiction reprend les éventuelles modifications des modalités de la sélection des faits réels employés dans la lecture de la fiction, lesquelles ont été mises en évidence dans §1.4.1. De même que dans le cas des États Men-

¹¹⁸Par exemple, Baddeley (1992, p12) affirme que son administrateur central est le « système attentionnel de contrôle qui est responsable pour la sélection des stratégies et pour l'intégration de l'information d'autres sources ». Schacter (1989), Baars (1988) et Johnson-Laird (1983, Ch. 16) sont encore plus explicites sur le rapport entre le « système directorial » et l'attention ou la conscience.

¹¹⁹Selon Schacter (1989), le système phénoménologique est une « porte » au système directorial.

¹²⁰Par exemple, selon Cowan (1993, p162), l'activation est « un état temporaire des représentations de mémoire qui permettent des effets d'amorçage de ces représentations sur des stimuli ultérieurs » ; ou selon Anderson (1983, p87), « l'activation est la mesure de la probabilité qu'un élément particulier de la connaissance soit utile à un moment particulier ».

¹²¹Cowan, *ibid.* et 1988, où Cowan soutient également un rapport important entre le *focus attentionnel* et le système directorial.

taux discutés auparavant, les lectures qui mettent en œuvre des « modes de sélection » différents (§1.4.1) semblent requérir des différences dans les détails des mécanismes psychologiques. Il faudrait, d'un côté, des différences dans le *genre* de simulation en jeu, dans la mesure où deux « simulations » relatives à deux lectures différentes peuvent s'appuyer sur des « ressemblances » différentes avec les processus « ordinaires ». Il y aurait, de l'autre côté, des rapports différents entre les entités mentales du statut « fictionnel » et les entités « ordinaires » selon les lectures différentes. Dans la mesure où l'essentiel du travail de la mise en rapport des aspects mentaux « fictionnels » et « réels » appartient non pas aux mécanismes proposés mais à un *autre* mécanisme, la question des différences ou des changements dans les modalités de la fiction entre des lectures différentes est une question des différences et des changements dans certains « comportements » de cet autre mécanisme (opérations appropriées du système directorial ; aspects de la dynamique de l'activation). De même que la théorie du *faire comme si*, ces théories des mécanismes psychologiques ne permettent pas de comprendre par eux-mêmes les modifications éventuelles dans les modalités de la lecture de la fiction ; il leur faut un fort appui sur quelque mécanisme externe, rarement explicité, et souvent lui-même difficile.

En somme, on ne saurait pas être totalement satisfait avec la réponse offerte par les théories proposant des mécanismes psychologiques à la question détaillée de la fiction, *dans la mesure où* elles ne permettent pas d'aborder ses complexités, les reportant plutôt à d'autres mécanismes qui ne sont pas encore explicités ni précisés. Ce qui vaut *et bien* pour les rapports complexes à la réalité relatifs à un instant *et bien* pour les expériences phénoménales du lecteur à l'égard de ces rapports *et bien* pour les éventuelles modifications des modalités de lecture de la fiction.

Pour reprendre la métaphore employée dans §1.4.1, ces théories travaillent en termes des vitesses moyennes, là où il faut quelque chose de plus spécifique, comme la vitesse instantanée. Ils répondent aux intuitions, mais non pas aux questions.

1.5.3 Limites des mécanismes usuels

Où l'on achève la discussion des approches psychologiques proposées dans la littérature en constatant que les théories des mécanismes « usuels » rencontrent des difficultés semblables à celles que l'on a mises en évidence relativement aux mécanismes spécifiques à la fiction. En outre, de même qu'auparavant, elles abordent ces difficultés

en les référant à un mécanisme « central ». Les problèmes apparemment communs à toutes les approches de la fiction semblent appeler une nouvelle stratégie pour aborder la question de la psychologie de la fiction ; en outre, vu que ces problèmes se posent également pour les théories établies des mécanismes ou des États Mentaux usuels, cette stratégie n'est pas contrainte à respecter ces théories.

Les théories des États Mentaux spécifiques à la fiction et les théories des États Mentaux usuels se heurtent aux mêmes difficultés (§1.4), à savoir, l'incapacité d'explicitement correctement l'articulation entre différents États Mentaux à des moments particuliers, de comprendre le rapport entre les États Mentaux et la phénoménologie du sujet à des moments particuliers, et de rendre compte des éventuelles modifications des modalités des États Mentaux. Cette communauté de défauts suggère que la difficulté n'est pas entièrement imputable aux à la théorie des États Mentaux spécifiques, mais a trait à l'usage des États Mentaux pour aborder la question détaillée de la fiction. On vient de montrer, dans §1.5.2, que les mécanismes proposés comme spécifiques à la fiction (entités mentales de statut « fictionnel », processus « fictionnels » de simulation) connaissent des difficultés tout à fait semblables. Pour terminer cette discussion des approches de la question psychologique de la fiction qui ont été proposées dans la littérature, on remarque que de semblables difficultés touchent également aux mécanismes « usuels », ce qui permettra une approche de la fiction qui, si elle s'oppose aux approches relatives à la fiction discutées jusqu'ici, n'est pas tenue de respecter les théories générales du mental dont ces approches se sont inspirées. En fait, il s'agit des trois phénomènes suivants :

1. la relation entre différents mécanismes « usuels » (différentes entités mentales, différents processus) n'est souvent pas comprise, surtout à des moments importants où ces mécanismes entrent en jeu, sans recourir à d'autres mécanismes « centraux » (système directorial, niveau d'activation).
2. la relation entre ces mécanismes et la phénoménologie de l'expérience du lecteur est également mal comprise, à ces mêmes moments.
3. On comprend à peine comment le statisme induit par les structures fixes de ces mécanismes permet de rendre compte des possibles modifications dans leurs modalités de mise en jeu.

Pour mettre en évidence les difficultés posées aux théories des mécanismes usuels, il suffit de transcrire les phénomènes et les exemples relatifs aux États Mentaux usuels (§1.4.2) sur le plan des mécanismes. Plutôt que de répéter les arguments présentés déjà trois fois, on se contente de considérer un seul exemple,

ayant trait au rapport entre différents mécanismes (la première difficulté), en l'occurrence l'exemple du rapport entre votre attitude à l'égard de la couleur des yeux d'un ancien camarade de classe et vos attitudes envers l'argument du paragraphe précédent au moment où vous le lisez. Dans §1.4.2, en discutant cet exemple, on a qualifié la couleur des yeux de votre ancien camarade de *hors sujet*. Sur le plan de mécanismes, on admet communément qu'il y a quelque entité mentale qui correspond à la croyance à l'égard de la couleur des yeux, et également qu'il y a des entités mentales correspondant aux aspects et attitudes qui sont en jeu actuellement ; la question est donc celle du rapport entre elles. Pour aborder cette question, il faut, comme dans le cas des mécanismes spécifiques à la fiction discutés dans le paragraphe précédent, recourir tantôt à quelque système central, tel le dit système directorial, tantôt à la notion d'activation. L'entité mentale relative à la couleur des yeux est *hors sujet* tantôt dans la mesure où le système directorial « ne s'occupe pas d'elle » à ce moment, tantôt dans la mesure où son niveau d'activation est nul (ou du moins, très bas), alors que les autres entités mentales « en jeu » sont actuellement « manipulées » par le système directorial, ou actuellement munis d'un niveau d'activation non nul (ou suffisamment élevé). D'un autre côté, il n'est pas difficile de voir que la discussion et les arguments proposés dans le cas de la fiction (§1.5.2), visant à montrer que la phénoménologie du sujet et les modifications des modalités de ses attitudes ne peuvent être comprises sans recours à des mécanismes « centraux », s'appliquent également dans le cas « ordinaire ».

L'importance de tels « mécanismes centraux » est d'ailleurs bien connue et communément admise, surtout lorsqu'il s'agit des capacités humaines de « niveau supérieur », telles le raisonnement ou la planification de l'action. Par exemple, beaucoup ont soutenu que, alors que certaines capacités mentales, telle la vision, peuvent être pensées comme effectuées par des « modules » dotés de capacités suffisamment « limitées », d'autres capacités, tels le raisonnement ou l'inférence, qui mobilisent la possibilité de référer à toute une gamme d'autres aspects psychologiques (visuelles, mémorielles, auditives, et ainsi de suite) dépendent fortement des mécanismes ou systèmes « centraux »¹²². Rétrospectivement, le fait pour les mécanismes fictionnels de dépendre de ces mécanismes « centraux » peut être rapproché du fait que la compétence fictionnelle compte sans doute parmi les capacités mentales de « niveau supérieur ». En outre, comme déjà remarqué dans §1.5.2, il faut non seulement comprendre

¹²²Le caractère spécialisé et, dans ce sens, limité des modules est explicité par Fodor (1985, Ch. III) et Baars (1988, §§1.4 et 2.1), par exemple. Voir Fodor, *op. cit.*, Ch. IV, ou Baars, *op. cit.*, Ch 2, sur la nécessité du recours à un « système central » et quelques remarques à son égard.

quels rapports ces mécanismes « centraux » (système directorial, partition d'activation) établissent entre *quels* mécanismes (entités mentales, processus ou modules), mais comment ces mécanismes « centraux » gèrent la *dynamique* de ces rapports, ce qui n'est aucunement une question facile¹²³. À tel point que certains ont même exprimé des doutes sur la possibilité de résoudre la question¹²⁴.

On est en position de résumer cette longue discussion. Aussi bien les approches « conceptuelles » que les approches « causales » à l'égard de la fiction adoptent la stratégie « mimétique » (§1.2.1) : elles posent une distinction stricte entre un aspect mental fictionnel et un aspect mental réel (*faire comme si* et croyances, entités mentales de statut fictionnel et ordinaire, simulations et processus ordinaires), pour insister après sur certaines similitudes entre les termes distingués. Dans tous les cas, ces approches ont peine à rendre compte de trois phénomènes – le rapport complexe entre attitudes « fictionnelles » et « usuelles » à des moments particuliers, le rapport complexe entre ces attitudes et la phénoménologie du lecteur à ces moments, les éventuelles modifications des modalités de ces attitudes. De plus, dans la mesure où l'expérience phénoménale du lecteur paraît correspondre aux rapports complexes entre les attitudes, les deux premiers phénomènes sont reliés : en échouant à rendre compte de l'un d'entre eux, les théories de la fiction échouent par là-même à rendre compte de l'autre. Par ailleurs, on trouve des phénomènes semblables relatifs aux attitudes « usuelles », lesquels sont plus ou moins problématiques pour les théories « conceptuelles » et « causales » à leur égard. Quant aux États Mentaux, on tire la conclusion qu'ils ne conviennent pas pour aborder les questions détaillées, telle celle de la fiction. Quant aux mécanismes psychologiques, on tire la conclusion que ce sont des mécanismes « centraux », plutôt que les mécanismes spécifiques proposés, qui tiennent la clef de ces questions détaillées. Or, malgré les nombreux

¹²³En règle générale, l'aspect dynamique est souvent peu compris sinon ignoré. Pour ceux qui parlent des systèmes directoriaux, il y a assez peu de résultats précis. Par ailleurs, les théories spécifiquement consacrées au raisonnement n'offrent que peu sur la question dynamique : les modèles mentaux proposés par Johnson-Laird (1983) semblent requérir quelque chose pour les gérer dans la dynamique, qui n'est pas explicité dans le détail ; les théories des changements de croyances de Stalnaker (1984) ou de Gärdenfors (2000) impliquent des paramètres de « similitude » ou de « saillance » respectivement, dont les valeurs, et notamment les dynamiques, sont généralement peu compris. Seuls les théoriciens du niveau d'activation ont suggéré des principes dynamiques précis (Anderson (1983, p92-97), par exemple), mais, même faisant abstraction des paramètres inconnus, on pourrait douter qu'ils offrent une véritable *compréhension* des capacités de « niveau supérieur » ; voir §1.5.2, notamment la note 116.

¹²⁴Fodor (1985, p140) : « plus un processus cognitif est global (c'est-à-dire isotrope), moins on sait des choses à son sujet. Sur les processus *très* globaux, comme le raisonnement par analogie, on ne sait rien du tout. » Pour un avis contraire, voir par exemple (Holland et al., 1989, notamment p335 *sq.*).

progrès, on n'est pas encore arrivé à comprendre les complexités psychologies relatives à la fiction en ces termes ...

Si l'on ne sait pas se contenter avec les réponses offertes à la question psychologique de la fiction, dont toutes empruntent la stratégie que l'on a qualifiée de « mimétique », il est peut-être temps d'en essayer un autre. Comme on verra dans le paragraphe prochain, il y a de bonnes raisons à croire qu'une stratégie de la sorte esquissée dans §1.2.3 arrivera à surmonter les difficultés mises en évidence dans la discussion précédente.

1.5.4 Repartir de la conscience

Où l'on propose et présente la stratégie qui sera empruntée dans les chapitres suivants. Les théories considérées précédemment posent d'abord une distinction entre des aspects mentaux (États Mentaux, mécanismes) « fictionnels » et « usuels », mais rencontrent ensuite des problèmes dans les rapports entre ces aspects et dans les expériences phénoménales du lecteur. Par opposition, la stratégie esquissée ici, qui est déjà anticipée dans §1.2.3, prend comme point de départ des expériences phénoménales, en étudiant directement les états de la conscience et leur dynamique, et en comprenant la fiction et le rapport à la réalité sur ce plan. Par ailleurs, des États Mentaux peuvent être attribués au sujet à partir de la succession de ses états de la conscience, en sorte que le rapport entre ses États et l'expérience phénoménologique est donné dans l'attribution.

La problématique développée et raffinée dans ces deux sections, qui constitue un vrai défi pour les approches de la psychologie de la fiction déjà proposées, se trouve complètement renversée dans la perspective « internaliste » adoptée dans ce travail (§1.1.2). Car, dans cette perspective, la phénoménologie du lecteur, laquelle pose tant de difficulté pour les approches considérées ci-dessus, est prise comme *donnée*. Il n'est donc plus besoin de poser des distinctions tranchantes entre États Mentaux ou mécanismes « fictionnels » et « usuels » pour ensuite se heurter à la question de leurs rapports à la phénoménologie du lecteur, ou d'ailleurs à la question reliée de leurs rapports entre eux. Il est également possible de *commencer* par la phénoménologie, pour venir *ensuite* à la question des distinctions possibles et convenables entre aspects fictionnels et réels. C'est la stratégie qui sera empruntée dans les chapitres suivants pour développer une théorie psychologique de la fiction ; le but du paragraphe actuel est de la présenter et de la motiver.

La différence fondamentale par rapport aux théories considérées ci-dessus, laquelle permet la conception d'une nouvelle stratégie pour aborder la fiction, concerne les données admises. Alors que les théories considérées ci-dessus adoptent l'externalisme méthodologique, selon lequel les données de l'enquête psychologique sont toutes manifestes (§1.3.2), la perspective « internaliste » adoptée ici (§1.1.2) prend les expériences *phénoménologiques immédiates* du lecteur, ou autrement dit ses *états de la conscience*. L'*état de la conscience* est à entendre dans le sens le plus large, comportant *tout ce qui est dans la conscience* : non seulement, pour parler comme Husserl, les objets, les pensées ou les images « vers lesquels la conscience est tournée » (par exemple ces mots sur la page), mais ceux qui sont dans « l'arrière-plan » (les détails de la page, les connotations des mots, et ainsi de suite) ; non seulement, pour parler comme James, ceux qui sont « au centre » mais « le halo ou le pénombre » qui est « en marge » ; non seulement, pour ainsi dire, ce qui est *focalisé* mais également ce qui ne l'est pas¹²⁵. Force est de souligner l'étendue de l'état de la conscience dans cette acception : dans la mesure où elle comporte « la marge aussi bien que le centre », elle déborde notamment ce que les psychologues cognitivistes discutent sous le nom « d'expérience phénoménale » ou de « conscience » (*awareness*). De fait, elle se rapproche à plusieurs égards de leurs notions plus techniques d'attention, telle l'attention comme degré d'activation signifiant (§1.5.2). Par exemple, les « informations » qui étaient sujet de l'attention dans le passé immédiat sont parfois relevées comme exemples typiques des informations « activées » mais hors de l'attention phénoménologique¹²⁶, alors qu'elles sont toujours « dans la conscience », quoique non pas « à son centre », comme le constatent Husserl et James¹²⁷.

Le fait de considérer les états de la conscience (pris dans cette acception large) comme des données fait apparaître sous un nouveau jour les problèmes qu'ont rencontrés les théories psychologiques de la fiction. Dans cette optique, les deux premières difficultés touchant aux approches « conceptuelles » aussi bien que « causales » (1. et 2. de §§1.4.1, 1.4.2 1.5.2, et 1.5.3) s'avèrent ramener à la question du rapport entre les aspects mentaux « fictionnels », les aspects mentaux « usuels » et les *états de la conscience*. Ceci est évident pour

¹²⁵James (1981, p255), Husserl (1913a, p62), par exemple. On mentionne ces noms *seulement* pour donner une idée de l'étendue de la notion de conscience en œuvre ici. Il ne s'agit de souscrire ici à aucune de leurs thèses relatives à la conscience.

¹²⁶Cowan (1993, p162). Naturellement, pour dialoguer avec ces théories psychologiques, il faut considérer ces « informations » comme des « représentations » du modèle computationnel de l'esprit qui sont accessibles ou « actives ».

¹²⁷La « mémoire primaire » de James, *op cit.*, Ch. 16 ; la « rétention » de Husserl (1928, §11 *sq.*).

la difficulté 2., qui concerne directement le rapport entre les aspects mentaux « fictionnels » (les *faire comme si*, les entités mentales de statut fictionnel, les simulations), les aspects mentaux « usuels » (les croyances, les entités mentales de statut ordinaire, les processus ordinaires), et les expériences phénoménales du lecteur – c'est-à-dire, ses *états de la conscience* – à des moments particuliers. Quant à la première difficulté (1.), ayant trait aux rapports entre les aspects mentaux « fictionnels » et « usuels » à ces moments, on a remarqué dans §§1.4.1, 1.5.2, et 1.5.3 que ces rapports correspondent apparemment aux aspects de la phénoménologie qui constituent la difficulté 2. Ce qui donne à croire qu'une résolution de cette dernière difficulté fournirait une résolution à la première. D'ailleurs, les deux difficultés, en tant qu'elles se posent pour les mécanismes, se rabattent l'une sur l'autre : car, le rapport entre les mécanismes « usuels » et les mécanismes « spécifiques à la fiction » est censé être géré par un mécanisme « central » (système directorial, partition d'activation), qui a, pour aspects « pertinents » (entités « en jeu » pour le système directorial, entités dotées d'un niveau d'activation significatif), ceux qui sont couverts par l'état de la conscience (entendu dans son sens large). Le problème commun à toutes ces théories est donc bien celui du rapport entre les aspects mentaux « fictionnels », les aspects mentaux « ordinaires » et les états de la conscience. On peut résumer cette problématique de manière schématique et simplifiée comme suit. Pour un instant t , écrivons B_t pour les croyances authentiques, respectivement les entités mentales « ordinaires » d'un mécanisme, F_t pour les croyances spécifiques à la fiction, respectivement les entités mentales « fictionnelles »¹²⁸ et C_t pour l'état de la conscience à l'instant t . La question est celle du triplet (B_t, F_t, C_t) et de sa dynamique (rapport entre (B_t, F_t, C_t) et $(B_{t'}, F_{t'}, C_{t'})$ pour t et t' différents)¹²⁹.

Or, dès que l'on présente le problème de cette manière, les approches traditionnelles de la question, qui commencent par poser les B_t et F_t (sur la base des stimuli et des comportements du sujet observés à la troisième personne) pour ensuite essayer de retrouver les C_t et le rapport entre les trois, perdent une large part de leur intuition et de leur intérêt. Plus précisément, du point de vue « internaliste », la stratégie « mimétique » (§1.2.1) adoptée par ces approches est soit *ad hoc* soit *maladroite*. *Ad hoc* : puisque les stimuli et les comportements observés à la troisième personne ne sont pas admis comme données dans cette perspective, la stratégie pose un certain nombre d'États ou d'entités mentaux (les B_t et les F_t) d'une manière qui est, du point de vue « internaliste », peu

¹²⁸On simplifie en considérant une distinction entre les mécanismes selon le statut des entités mentales ; cette simplification ne touche cependant pas à la généralité des conclusions, qui s'appliquent également aux autres mécanismes psychologiques proposés.

¹²⁹Bien entendu, on simplifie la question quelque peu, en omettant par exemple les désirs.

comprise. *Maladroite* : on pourrait néanmoins « attribuer » les B_t et les F_t dans la perspective « internaliste », non pas sur la base des stimuli et des comportements observés à la troisième personne, mais sur la base des stimuli et des comportements *observés à la première personne*. Or, cette manière de procéder est déconcertant : elle consiste à poser les B_t et les F_t sur les bases de *certain*s aspects des C_t , pour ensuite essayer d'aborder le rapport difficile entre les B_t , les F_t , et *d'autres* aspects des états de la conscience C_t . Bref, non seulement le choix de partir des croyances et des « *faire comme si* » (ou des aspects mentaux usuels et des aspects mentaux spécifiques à la fiction) n'a pas déjà porté des fruits en pratique, il est suspect sur le plan théorique aussi.

Une autre stratégie s'impose : au lieu de partir des croyances et des faire comme si (les B_t et les F_t), partez des états de la conscience (les C_t).

Il importe de désamorcer tout de suite une éventuelle objection contre cette stratégie, à savoir qu'elle fait fi des théories des États Mentaux et des mécanismes usuels qui ont été déjà développées, c'est-à-dire des théories s'occupant de la paire (B_t, C_t) et de sa dynamique, et des éventuels apports de ces théories et de leurs pour la question de la fiction. Cette objection ne tient pas, car, comme l'on a montré dans §§1.4.2 et 1.5.3, ces théories souffrent des difficultés apparentées à celles qui ont touchées aux theories des États Mentaux et mécanismes spécifiques à la fiction. En effet, grâce à la communauté des difficultés qui appartiennent aux questions de (B_t, C_t) et de (B_t, F_t, C_t) , rien n'empêche de renverser *les deux questions*, en abordant les deux à partir de la conscience C_t ¹³⁰.

Il est cependant vrai que les États Mentaux et les mécanismes ont désormais une importance amoindrie. Car, tout simplement, *la fiction et son rapport à la réalité* (dans la psychologie du lecteur) *peuvent être abordés directement en termes de la conscience, sans faisant entrer ses États Mentaux ou des mécanismes*. D'un côté, les données du problème (en tant que spécifiées dans §1.1) sont les états de la conscience C_t , de telle sorte que la question du rapport entre les C_t et les $C_{t'}$ a bien un sens indépendant et « primitif », et *pourrait* être abordée sans mettre en jeu la question des B_t et des F_t . De l'autre côté, dans la mesure où les pensées, les images, et les affects relatifs à la fiction, et d'ailleurs pertinents pour son rapport à la réalité, apparaissent et se montrent de manière plus ou moins explicites dans les états de la conscience, la question du rapport entre fiction et réalité et de sa dynamique *pourrait* être posée en termes des états de la conscience *seuls*, sans s'appuyer sur ou mettre en jeu les États Mentaux ou d'autres *posits* « fictionnels », tels le *faire comme si* ou la simulation.

¹³⁰On se concentrera tout de même sur la question relative à la fiction dans la suite.

Néanmoins, il ne découle pas de cette autonomie de la conscience que l'on ne peut pas retrouver les États Mentaux, telles la croyance ou éventuellement le *faire comme si*¹³¹. Car, alors que l'on attribue des États Mentaux typiquement sur la base des comportements du sujet face à certains stimuli, on pourrait faire le même *genre* d'attribution sur la base de la succession de ses états de la conscience. Autrement dit, de même que les États Mentaux de la tradition sont attribués dans des *blocs d'attribution* qui comportent des comportements du sujet dans certaines situations (§1.3.2), on pourrait toujours attribuer des États Mentaux dans le cadre « internaliste », par le biais des blocs de structure semblable qui figurent, plutôt que des comportements du sujet, ses *états de la conscience*.¹³²

La stratégie proposée prend une perspective différente non seulement sur les aspects psychologiques relatifs à la fiction, mais sur l'approche de la fiction en général : en effet, elle implique un rejet de la stratégie « mimétique » qui est empruntée habituellement (§1.2.1). À sa place, elle propose d'aborder *directement* l'entière d'un champ psychologique où la distinction entre fiction et réalité *est a priori absente* (en l'occurrence, le champ de la conscience) pour dissocier, *dans un deuxième temps*, aussi bien les aspects divergents relatifs à la fiction et à la réalité que leurs aspects communs (dont éventuellement les différents États Mentaux spécifiques à la fiction et à la réalité). Il s'agit donc de la stratégie « naturaliste » esquissée dans §1.2.3 : au lieu de *poser* la distinction entre fiction et réalité en entrée de jeu, avec ou par la distinction entre le *faire comme si* et la croyance, on la *retrouve* après coup comme des aspects, des modalités et des propriétés des états de la conscience et de leur dynamique.

Pour résumer grossièrement, la stratégie pour aborder le triplet (B_t, F_t, C_t) comporte *deux moments*. Premièrement, on se donne les états de la conscience C_t , et aborde la question de la dynamique de ces états (approximativement, le

¹³¹Les États Mentaux sont des concepts sociaux mobilisés pour « comprendre » le comportement de l'autrui, et donc il est important de montrer comment on les retrouve dans le cadre proposé ; en revanche, puisque les mécanismes sont des aspects des différentes théories psychologiques plus ou moins techniques, la question du rapport entre ces mécanismes et le cadre proposé est moins importante.

¹³²Bien entendu, si l'on considère que le contenu d'un État Mental (la proposition que l'on croit, par exemple) est déterminé (partiellement) par des aspects sociaux ou externes au sujet, comme le soutient la thèse « l'externaliste » en philosophie de l'esprit (Putnam, 1975; Burge, 1979), l'attribution décrite ici n'arrive pas à fournir un contenu de ce genre. Il n'est pourtant pas question ici de la conception appropriée du contenu des États Mentaux, mais plutôt du fait que le choix de travailler à partir de la conscience n'empêche pas l'attribution des États Mentaux (ou, du moins, de quelque chose qui y ressemble fortement), et d'ailleurs des États Mentaux ayant des propriétés familières et répandues. Voir également la note 59 de §2.4.

rapport entre C_t et $C_{t'}$ pour t et t' différents). Une compréhension du rapport entre fiction et réalité et de la dynamique de ce rapport est déjà disponible à ce stade. Deuxièmement, à partir des états de la conscience C_t , on attribue les États Mentaux de croyance et de *faire comme si* (B_t et F_t). Ce stade est requis seulement pour mettre l'analyse de la conscience en rapport avec les concepts mentaux employés quotidiennement pour parler de la fiction et des attitudes à son égard. Ces deux étapes permettent une théorie de (B_t, F_t, C_t) et de sa dynamique : d'un côté, le rapport entre B_t , F_t , et C_t à un moment t particulier est déjà spécifié par l'attribution, de l'autre côté, sa dynamique est abordée en mobilisant la réponse à la question de la dynamique des C_t et les modalités de l'attribution des B_t et F_t .

Dans les deux chapitres suivants, on s'efforcera de développer une théorie du rapport entre le fiction et le réalité dans la psychologie du lecteur qui adopte cette stratégie. Conformément à la difficulté de la question posée, on ne parviendra qu'à poser les éléments de base d'une telle théorie, lesquelles ne fournissent pas déjà une réponse complète et pleine à la question de la fiction. Néanmoins, ils suffisent pour rendre compte d'un grand nombre de phénomènes relatifs à la fiction, pour apercevoir d'autres phénomènes sous un nouveau jour, et pour offrir des pistes prometteuses envers une théorie approfondie.

1.6 Résumé : la question de la fiction et ses stratégies

La question générale qui domine la première partie de cette thèse est celle du rapport entre fiction et réalité, en tant qu'elle se pose relativement à la psychologie du lecteur.

Les quatre questions posées au début du présent chapitre (§1.1.1) ont trait à ce rapport. Chaque question témoigne, à sa façon, d'une *distinction cruciale* entre fiction et réalité, mais également d'une certaine *similitude* entre elles.

Or, le simple constat d'une distinction et d'une similitude ne suffit pas pour comprendre à fond le rapport entre fiction et réalité : il faut plutôt préciser *en quoi* la fiction se distingue de la réalité et *en quoi* elles se ressemblent. *A priori* cette question n'est pas facile ; en effet, elle *constitue* le noyau du problème de la fiction (§1.1.3).

Une stratégie à l'égard de la question du rapport entre fiction et réalité – la stratégie « mimétique » – prend comme point de départ ces deux aspects apparents du rapport. Plus précisément, elle pose d'abord la distinction entre

fiction et réalité, en affirmant que la fiction n'est pas réelle, pour ensuite restituer d'une manière ou d'une autre une similitude entre elles (§1.2.1).

Sans explicitation du sens de cette distinction et de cette ressemblance, cette stratégie ne fait que répéter les données à expliquer. Néanmoins, il se peut que la méthode qui consiste à poser une telle distinction comme *petitio principii* soit fructueuse, permettant, par le biais du développement de cette distinction et de la ressemblance associée, d'atteindre une compréhension raffinée des subtilités du rapport entre la fiction et la réalité.

Pourtant, cette stratégie, en tant qu'on l'adopte pour répondre à la question relative à la psychologie du lecteur, mène dans l'embarras. En particulier, les théories empruntant cette stratégie ont peine à rendre compte, premièrement, des complexités du rapport entre fiction et réalité (en tant qu'aspects psychologiques) à plusieurs instants de la lecture ; deuxièmement, des complexités du rapport entre les aspects mentaux « fictionnels », les aspects « réels », et la phénoménologie du lecteur à ces instants ; et dernièrement de l'éventualité des variations dans les modalités de la distinction et de la similitude entre fiction et réalité, selon l'époque, l'œuvre et la lecture (§§1.4-1.5).

De fait, il s'agit pour l'essentiel des instants, et plus généralement des aspects, qui sont cruciaux pour répondre aux quatre questions posées dans §1.1.1 (instants où la compétence fictionnelle du lecteur est mise à l'épreuve, où surgissent les affects relatifs à la fiction, où la réalité entre en jeu dans la lecture, et ainsi de suite). Ainsi, les théories proposées, lesquelles empruntent la stratégie « mimétique », n'offrent pas de réponses satisfaisantes à la question de la fiction.

Il est donc besoin d'adopter une autre stratégie et de prendre un autre point de départ. Les considérations précédentes en suggère déjà une, laquelle ne pose pas la distinction entre la fiction et la réalité comme une distinction entre deux types d'États Mentaux ou de mécanismes, mais la retrouve à partir d'un niveau psychologique où elle *n'est pas déjà présente* d'une telle manière tranchante (§1.2.3). Ce niveau n'est pas celui des États Mentaux, ni des mécanismes traditionnels, mais celui de la conscience (§1.5.4). Le rapport avec la conscience est embarrassant pour les théories traditionnelles ; en le prenant comme point de départ, on s'attend à éviter ces problèmes.

Le reste de cette partie est consacré à développer une théorie psychologique *précise* de la lecture selon cette stratégie. Plus précisément, dans le Chapitre 2, on ébauchera une théorie générale du mental qui est fondée sur la conscience. Cette théorie servira, dans le Chapitre 3, à porter des éléments de réponse aux quatre questions posées au début de la thèse, et ainsi à la question générale du rapport entre fiction et réalité.

Chapitre 2

La psychologie du complexe : jeter les bases

Les approches traditionnelles de la question du rapport entre fiction et réalité dans la psychologie du lecteur ont peine à comprendre la dynamique de ce rapport. En effet, comme on l'a montré dans §§1.4.1 et 1.5.2, ces approches n'ont pas su comprendre les complexités du rapport entre fiction et réalité à des moments particuliers, ce qui ne peut que leur encore embarrasser relativement à la question de la dynamique de ce rapport, laquelle a été reconnue comme une partie majeure de la question générale de ce rapport dans la psychologie du lecteur. Or, ce manque n'est pas limité aux théories ayant trait aux États Mentaux ou aux mécanismes mentaux spécifiques à la fiction : les théories des États Mentaux ou des mécanismes « usuels » ont également peine à rendre compte des subtilités des rapports entre des croyances (par exemple) à un moment donné, ainsi que des complexités de la dynamique des croyances, ou plus généralement du mental (§§1.4.2 et 1.5.3). Qu'il s'agisse de la fiction en particulier ou du mental en général, une conception insuffisante de l'état instantané et une conception insuffisante de la dynamique constituent apparemment deux insuffisances liées.

Il semble donc nécessaire, avant de venir à la question spécifique du rapport entre fiction et réalité dans la psychologie du lecteur, de s'attarder sur des questions générales à l'égard du mental, et notamment à l'égard de son état instantané et de sa dynamique. La stratégie que l'on adopte ici pour aborder la fiction, laquelle a été esquissée dans §§1.2.3 et 1.5.4, commence d'ailleurs par travailler à un « niveau » mental où la distinction entre fiction et réalité n'apparaît pas d'abord, pour la retrouver après coup. Le but du chapitre actuel est de jeter les éléments de base d'une approche du mental qui permettra, dans le

chapitre suivant, de proposer une analyse des phénomènes relevant de la lecture de la fiction.

La tâche fondamentale préalable à toute théorie du mental est, on vient de la souligner, de *concevoir l'état instantané du mental de telle manière qu'elle permette une pensée féconde de sa dynamique*. L'approche proposée vise directement à relever ce défi, en prenant comme « unité de base » l'état de la conscience. Il faut cependant un travail précis et soigneux pour *bien* le concevoir, c'est-à-dire, pour concevoir l'état instantané de la conscience de telle manière qu'il étiât une pensée féconde de sa dynamique. Les trois premières sections de ce chapitre sont consacrées à la construction d'une telle conception.

La distinction entre fiction et réalité n'est pas donnée immédiatement au niveau de la conscience. Néanmoins, comme anticipé dans §§1.2.3 et 1.5.4, et comme on verra dans le Chapitre 3, on peut retrouver les aspects de la fiction et de son rapport à la réalité comme des aspects *dynamiques* de la conscience, comme des propriétés des *successions* des états instantanés de la conscience, qui sont associés à, ou sont des conséquences de, la connaissance que le lecteur a du fait que le texte est fictionnel, connaissance admise comme hypothèse de départ (§1.1.2). En un mot, *la question de la fiction et de son rapport à la réalité*, en tant qu'elle se pose relativement à la psychologie du lecteur, *trouve sa réponse au niveau de la conscience, des états instantanés de la conscience et de la dynamique de ces états*.

Il s'ensuit que, à strictement parler, les États Mentaux ne sont pas requis pour répondre à la question du rapport entre fiction et réalité dans la psychologie du lecteur ; pourtant, ce travail n'a aucune prétention révisionniste à leur égard. Au contraire, des parties importantes de ce chapitre et du prochain sont consacrées à montrer la possibilité de retrouver les États Mentaux ordinaires dans le cadre mental développé. Cette volonté de retrouver les États Mentaux ne relève pas seulement de la reconnaissance de leur utilité quotidienne, mais également de l'objectif de diagnostiquer exactement les problèmes qui touchent aux théories qui posent des États Mentaux spécifiques à la fiction (tels le *faire comme si*). En exprimant les États Mentaux suggérés comme spécifiques à la fiction, aussi bien que les États Mentaux usuels, en termes du cadre proposé, on sera en position de mieux comprendre les détails et les difficultés des rapports entre eux. Ce travail est reporté au chapitre suivant (notamment §3.3) ; on prépare cette discussion avec une considération générale des États Mentaux en tant qu'ils sont compris dans le cadre proposé, c'est-à-dire en tant qu'ils sont compris à partir des états de la conscience. L'idée, qui a été déjà mentionnée dans §1.5.4, est simple : il s'agit, de même que les approches traditionnelles aux

États Mentaux (§1.3.2), d'attribuer des États Mentaux dans des blocs d'attribution, comportant, à la différence des approches traditionnelles (§1.3.2), les états de la conscience du lecteur. Les deux dernières sections de ce chapitre sont consacrées à cette question.

2.1 Faire face à la conscience

L'approche du mental, et finalement du rapport entre fiction et réalité dans la psychologie du lecteur, qui sera proposée ici est fondée sur la conscience. Or, il est notoire que la notion de conscience entraîne souvent des confusions, des imprécisions et des difficultés fâcheuses. Ce qui exige une réflexion sur la notion de conscience que l'on emploie et la représentation que l'on en fait. Dans un certain sens, cette section et les deux qui suivent s'occupent de la question de la représentation de la conscience, et des conséquences immédiates qui peuvent être tirées à son égard.

En effet, la représentation doit satisfaire à des conditions diverses. D'un côté, elle doit être relativement rigoureuse, tout en étant fidèle à l'état de la conscience à l'instant approprié. De l'autre côté, elle doit rendre compte de la question importante de la dynamique de la conscience. Car la prise en charge de la dynamique exige une représentation *adéquate* de la conscience, c'est-à-dire une représentation qui est féconde et fidèle non seulement par rapport à l'état instantané de la conscience mais également par rapport à sa dynamique. Il s'agit donc d'une exigence sur les *deux* aspects à la fois : d'une part, il faut concevoir la dynamique de la conscience en termes d'une représentation de l'état instantané de la conscience, et d'autre part, il faut construire une représentation de l'état instantané de telle sorte qu'elle permet d'étayer une conception féconde et fidèle de la dynamique. Dans un certain sens, l'erreur de ceux qui posent d'abord la distinction entre fiction et réalité (sur le plan d'aspects mentaux, par exemple) est d'ignorer la difficulté de penser la dynamique du rapport entre fiction et réalité en termes de cette opposition rigide.

La section actuelle a pour but d'introduire la notion de conscience dont il sera question dans la suite, et de proposer quelques outils techniques qui seront employés pour parler des états de la conscience à des instants particuliers avec précision. Dans la section suivante, on repère certaines propriétés immédiates et fondamentales de l'état de la conscience à un instant donné, qui permettent d'approfondir la représentation que l'on en fait, pour en venir, dans §2.3, à la question de la capacité de la représentation proposée à prendre en charge la dynamique de la conscience. De ce qui précède, il devrait être clair que l'ordre

suivi est plutôt un *ordre d'exposition* qu'un *ordre d'analyse*, car la question de l'état instantané et celle de la dynamique ne sont pas des questions séparées, qui se posent l'une après l'autre, mais plutôt des questions imbriquées, qui se posent ensemble.

Pour clarifier le *statut* de la théorie dont on a affaire dans ce chapitre, il convient de la comparer aux deux sortes des théories du mental discutées dans le chapitre précédent. De même que les théories proposant les États Mentaux (§1.3.1), le but est de développer un certain nombre de *concepts* permettant une *compréhension* du mental ; or, les concepts proposés ici sont plus *fins* que les notions de croyance et de *faire comme si* considérées ci-dessus. Sur le plan de la finesse donc, les concepts développés ici approchent des mécanismes proposés par les théories qui s'occupent des dites questions « *causales* » (§1.3.1) ; pourtant, à la différence de ces théories, il n'est aucunement question ici de la réalisation physique ou informatique des aspects de la théorie. Plutôt qu'un mécanisme qui pourrait (théoriquement) être construit ou un programme qui pourrait (théoriquement) être exécuté, il s'agit de proposer quelque chose de l'ordre d'une *analyse* qui pourrait (théoriquement) être effectuée. Analyse : non pas au sens (philosophique) d'une décomposition d'une notion en ses constituants, mais plutôt au sens (scientifique) d'un traitement d'un ensemble de données. Les données consistent en des états de la conscience à des moments successifs ; leur traitement appelle des concepts méticuleusement développés et subtils.

Dans le premier paragraphe de la présente section, on précise l'acception de la notion de conscience en tant qu'elle est employée dans cette thèse et de certaines notions liées qui seront utiles. Dans le second paragraphe, on pointe quelques problèmes bien connus qui se posent pour des théories qui s'appuient sur la conscience et on indique comment on les contourne ici. Dans le dernier paragraphe, on propose un certain nombre de concepts techniques qui permettent de discuter, voire de *modéliser* la conscience de manière rigoureuse et donc d'apporter un brin d'exactitude à la discussion ultérieure.

2.1.1 La conscience : intuitions et enjeux

Où l'on introduit la notion de conscience qui est en jeu dans ce travail, et on fixe le sens de « dans la conscience » qui est en vigueur.

Il s'agit d'aborder la vie mentale du lecteur, et finalement le rapport entre la fiction et la réalité en tant qu'il s'érige en elle, à partir de la *conscience*. Comme l'a été déjà souligné dans §1.5.4, « la conscience » est entendue dans ce

travail dans son sens le plus large, que l'on trouve par exemple chez James ou Husserl. Il comporte toute « expérience phénoménale », qu'elle soit en avant ou au fond, au centre ou en marge, au premier plan ou à l'arrière-plan, au point ou non. On aurait tort de sous-estimer *l'étendue* de la conscience ainsi conçue ; elle ne se limite pas notamment au *focus* de l'attention, comme on l'a souligné auparavant (§1.5.4). Il convient dans ce paragraphe de s'attarder sur d'autres aspects de la conscience en tant qu'elle se présente dans le cadre « internaliste » en vigueur ici. En premier lieu, on précise l'acception de quelques termes qui sont employés dans ce travail ; en dernier lieu, on souligne l'aspect de la conscience qui est crucial pour les propos de ce travail – sa dynamique – et on explicite la stratégie adoptée pour l'aborder.

Des philosophes parlent souvent de la conscience comme un « flux » (*stream of consciousness*) muni de « parties »¹. Alors que la métaphore du flux a le mérite de souligner l'importance de la dynamique de la conscience, elle est moins féconde sur la question des « parties » de la conscience. Pour expliciter ce qui compte comme « partie » de la conscience, en tant qu'elle se présente dans cette perspective « internaliste », il convient d'employer la métaphore qui rapproche la conscience du champ de vision². De prime abord, on voudrait dire que le « contenu » ou les « objets » de la conscience sont comme les objets dans son « champ de vision ». Naturellement, on frôle *normalement* des difficultés ici : on a tendance à dire que lorsqu'on regarde le Mont Blanc, c'est le mont lui-même qui est dans le champ de vision, mais lorsqu'on pense au Mont Blanc, est-ce qu'il est là (lui-même) dans la conscience ? Dans le cadre « internaliste » adopté pour ce travail, où seulement l'apparition au sujet est en jeu, la question du Mont Blanc (lui-même) et de sa présence est *hors sujet* ou *mise entre parenthèses*. Pour un observateur à la troisième personne qui peut comparer le champ de vision de l'œil (l'image rétinienne par exemple) au « vrai état du monde » devant l'œil, il y a deux sens possibles du « dans » dans l'énoncé « le Mont Blanc est dans le champ de vision » – l'un qui a trait à l'existence « effective » (du Mont Blanc) dans le monde devant l'œil, l'autre qui se réfère à une « présence » ou une apparition (de « l'image du Mont Blanc ») dans le champ de vision. En revanche, pour l'œil dont cette scène est le champ de vision, il n'en y a qu'un

¹James (1981, Ch IX), Husserl (1913b, V^e Recherche, Ch. 1, §§1 et 2), pour prendre deux exemples.

²On peut fusionner les deux métaphores, en se figurant la conscience comme le champ de vision à partir d'une barque qui descend le ruisseau (*stream*). Ceci aurait le mérite supplémentaire de souligner que les changements de la conscience sont *vécus* « dans la conscience » (comme les changements *dans* le champ de vision) plutôt qu'*observés* « sur la conscience » (comme on observe un ruisseau). Voir par exemple Wittgenstein (1922, §5.6331).

sens possible. En s'installant dans un « cadre internaliste », on se met dans la position de l'œil plutôt que de l'observateur, à telle enseigne qu'il *n'y a qu'un seul sens* du « dans » lorsqu'on dit « en pensant du Mont Blanc, le Mont Blanc est *dans* la conscience ». C'est le seul sens du « dans » dans l'expression « dans la conscience » (ou de « présent » dans « présent à la conscience »), expression que l'on emploiera fréquemment et sans équivoque dans la suite.

Naturellement, cet « objet dans » la conscience, du moins dans le cas de la perception, pourrait être rapproché de « l'objet (intentionnel) » de Husserl³, dont l'*epochè* a des ressemblances à l'« internalisme » adopté ici à l'entrée du jeu. Toutefois, l'étendue de son *epochè*, ainsi que d'autres détails de sa phénoménologie diffèrent de l'approche esquissée ici. En effet, comme on soulignera prochainement, il s'agit dans l'établissement et l'expression « précise » de la conscience, de faire entrer le minimum de thèses. On ne suppose donc *au départ* aucune de ses « analyses » des différents « aspects » de la conscience, aucune distinction entre actes et objets ou entre différents types d'actes, et ainsi de suite⁴. Si de telles distinctions sont soutenables, on les retrouvera « dans » la conscience – elles seront « présentes » dans les objets ou le contenu de la conscience et dans la structure de ces objets (en tant qu'ils sont présents dans la conscience).⁵

Si l'on préfère la métaphore du champ de vision à la métaphore du flux pour ce qui s'agit des objets « dans » la conscience, on ne saurait suffisamment souligner l'importance de la dynamique de la conscience qui est mise en avant par la métaphore du flux. La question de la fiction et de son rapport à la réalité *est* en large partie celle du dynamique de ce rapport, de sa modification et sa mutation pendant et après la lecture (§§1.4 et 1.5) ; l'importance de la conscience réside par conséquent dans *sa dynamique*. Alors que la discussion de cette section et de la prochaine porte pour l'essentiel sur *l'état instantané de la conscience* – l'état de la conscience à un instant particulier⁶ –, la question importante, à laquelle on en vient dans §2.3, est celle de la dynamique de l'état instantané de la conscience, c'est-à-dire de la succession d'états instantanés de la conscience, des relations entre un état instantané et les états ultérieurs et antérieurs.

³Husserl (1913a, §§39 et 41), par exemple ; ou §88, où il le nomme « noème ».

⁴On ne reconnaît donc pas la structure de « vécus » proposée par Husserl (1913a, par exemple §36 *sq.*).

⁵Voir par exemple la note 73 ci-dessous.

⁶La question de ce qui compte comme « un instant », et les questions ayant trait au temps en général sont évidemment importants pour toute théorie qui prend en charge la dynamique. Malheureusement, il n'est pas assez de place dans cet essai pour traiter ces questions.

2.1.2 Analyser l'état instantané de la conscience : introduction

Où l'on discute deux difficultés qui se posent pour une théorie qui prend comme données les états de la conscience. Premièrement, l'inaccessibilité de la conscience du sujet au théoricien pose une double question : celle, pratique, de l'acquisition des informations à l'égard de la conscience, et celle, méthodologique, de la perspective adoptée par le théoricien sur la conscience du sujet. Dans ce travail, où l'intérêt est théorique plutôt que pratique, on adopte la perspective (pratiquement intenable, mais théoriquement féconde) d'un théoricien qui « observe » les états de la conscience que le sujet « vit » de l'intérieur. Deuxièmement, se pose la question de la bonne manière de représenter ou décrire l'état de la conscience à un instant. La représentation est guidée par deux maximes, qui sont en tension : elle devrait être fidèle à ce qui est représenté, mais également féconde, dans la mesure où elle est convenable pour un travail théorique. Dans le prochain paragraphe, on proposera des concepts techniques qui sont aussi neutres que possible, permettant une représentation relativement fidèle de la conscience.

On s'efforce de concevoir une théorie qui prend comme « données » les états de la conscience d'un sujet – le lecteur⁷. Pourtant, une telle théorie bute tout de suite sur un certain nombre de difficultés bien connues, dont deux sont particulièrement épineuses. D'une part, la conscience de quelqu'un est ce qu'il éprouve, ou vit, à lui seul : certainement, il semble impossible en pratique pour un théoricien d'accéder à la conscience d'autrui (quoi que ça veuille dire). D'autre part, le nombre impressionnant de théories qui prétendent décrire ou s'appuyer sur la conscience mais qui introduisent de manière subreptice des présupposés douteux parmi leurs « observations » a servi à souligner les difficultés en *décrivant* les états de la conscience de manière fidèle. Ces difficultés sont d'ordre *méthodologique*, et donc appellent un approfondissement de la méthodologie en vigueur dans cette thèse. Le but du présent paragraphe est de clarifier les positions méthodologiques qui permettent de prendre en charge ces apparentes difficultés relevées.

La première difficulté, à savoir l'inaccessibilité apparente de la conscience

⁷Evidemment, bien que l'on travaille avec *une* collection des états de la conscience d'un sujet, les régularités auxquelles on s'intéresse et les conclusions auxquelles on aboutit sont généraux à tout sujet.

à l'autrui, et notamment au théoricien, comporte effectivement deux éléments. D'une part, il y a la question de l'*acquisition* des informations relatives à la conscience d'autrui – comment parvenir à recevoir ces informations ? D'autre part, il y a la question, sans doute reliée, de l'*approche* de la conscience d'autrui – quelle perspective prend-on sur sa conscience ? La sienne ? Celle d'une troisième personne ? Ces deux questions sont de caractère différent : la première est d'ordre *pratique*, alors que la dernière se pose aussi bien sur le plan *théorique* que pratique. Si bien qu'il est nécessaire de clarifier le *statut* de l'enquête menée dans cette thèse. Elle est *philosophique* ou *théorique*, cherchant à comprendre les propriétés générales du mental, plutôt qu'*expérimentale* ou *pratique*, visant à étudier directement des individus spécifiques à des moments particuliers.

Par conséquent, la difficulté d'acquisition ne se pose pas. En effet, une discussion théorique qui prend la perspective « internaliste », perspective où l'on ignore tout sauf le mental du lecteur, n'a pas de sens sauf si l'on suppose que tous les théoriciens parlent et « ont accès » aux mêmes expériences du lecteur ; en adoptant la perspective « internaliste » dans ce travail, on admet cette supposition. En revanche, la difficulté d'approche se pose toujours : quel genre de perspective adopte-t-on sur la conscience du sujet ? Est-elle, en particulier, celle que le sujet a sur sa propre conscience ? Non, car cette perspective limite les observations à celles que le sujet lui-même fait, et prive le théoricien de toute possibilité de repérer les phénomènes, telles les propriétés de la succession des états de la conscience, dont le sujet « ne tient pas compte ». En outre, si l'on prend la perspective du sujet, on risquera de confondre les expériences du sujet avec les observations du théoricien. Il vaut donc mieux distinguer la perspective du théoricien de celle du sujet, pour permettre au théoricien un certain recul, de sorte qu'il peut faire des constats à l'égard de la conscience du sujet dont le sujet n'est pas nécessairement « conscient ». La vue sur la conscience que l'on emploie ici n'est pas celle du sujet qui la *vit*⁸, mais celle d'un théoricien qui *l'observe la vivre*. Dans la mesure où le théoricien est distinct du sujet, il s'agit d'une vue à la troisième personne ; dans la mesure où il s'agit seulement des états de la conscience du sujet, c'est une vue à la première personne : on peut donc caractériser la perspective adoptée à l'égard de la conscience comme une vue à la troisième personne *sur* une vue à la première personne.

Bien entendu, cette perspective n'est nullement atteignable en pratique⁹. Il

⁸Le sujet ne pourrait pas quitter sa conscience pour la considérer de dehors. C'est le sens de l'affirmation selon laquelle le solipsisme est un réalisme (§1.1.2.) Voir Wittgenstein (1922, §5.62).

⁹On ne pourrait au mieux que s'approcher à cette vue (relative à soi-même) en s'efforçant de « s'éloigner » de sa propre conscience.

s'agit plutôt d'une *hypothèse fictive* : celle d'un sujet, dont les expériences et la conscience sont exposées devant un groupe de théoriciens qui discutent comment mieux comprendre ces données. Cette hypothèse est d'ordre *méthodologique* plutôt que *métaphysique*, *ontologique* ou autre. Il ne s'agit aucunement d'une *affirmation* relative à la nature du mental¹⁰. Au contraire, l'hypothèse a pour but de permettre le développement d'une théorie psychologique ; en l'occurrence, la théorie psychologique proposée ci-dessous se construit dans cette hypothèse. Par conséquent, elle devrait être justifiée en ces termes, c'est-à-dire par l'intérêt de la théorie que l'on peut formuler en s'appuyant sur elle.

Par ailleurs, la deuxième difficulté posée par la conscience concerne « l'expression » ou la « description » de la certaine organisation que compte un état instantané de la conscience. Selon l'hypothèse fictive adoptée, le théoricien *observe* la conscience du sujet, comme le physicien observe un oiseau en vol ; la difficulté est de la décrire *fidèlement*. Autrement dit : il lui faut trouver une manière *d'exprimer* ou (de représenter) ce que le sujet *vit* (ou éprouve). C'est la tâche, dont la complexité est par ailleurs bien connue dans la philosophie de sciences du moins depuis Kuhn¹¹, d'*établir les données* (de la théorie). Quant à cette tâche, il y a deux contraintes à respecter :

Fécondité : les données doivent être exprimées d'une manière convenable pour l'emploi dans et par une théorie intéressante.

Fidélité : les données doivent être exprimées d'une manière aussi proche de l'observation, et comprenant le moins de suppositions externes que possible.

Ces deux contraintes comportent des termes peu précisés. Dans la mesure où la fécondité à l'observation « brute » (« voilà un oiseau qui vole ») semble nuire à l'expression convenable pour des théories intéressantes (« voilà un corps avec vitesse v à point x »), et vice versa, les deux contraintes sont en tension. Aucun besoin cependant de s'attarder ici sur la précision ou la tension : la fécondité et la fidélité sont des *maximes* qui guident l'établissement ou la constitution des données pour la théorie.

Étant donné l'histoire de l'usage philosophique de la conscience, il vaut bien mettre l'accent sur la *fidélité* de la formulation des données, pour éviter d'intro-

¹⁰En particulier, aucune thèse n'est soutenue relativement à la « vraie » nature de la conscience et de sa représentation. Par conséquent, les arguments visant à montrer que tout énoncé du langage (ordinaire) faisant référence à la conscience est « dépourvu de sens » ne sont pas pertinents ici.

¹¹Kuhn (1970, p172) : « Étant donné un paradigme, l'interprétation des données est capitale pour qui se propose d'explorer ce domaine. » Voir plus généralement son Ch IX.

duire subrepticement des suppositions supplémentaires : de fait, on s'efforcera d'établir et d'employer des outils et des notions aussi *neutres* que possible pour respecter au plus ce maxime. Cependant, on ne pourrait pas négliger la maxime de *fécondité* : représenter la conscience sous forme d'une collection aléatoire et non structurée d'énoncés telles « il y a une table », « je suis heureux », « Holmes ne devrait pas faire ça » ... ne convient aucunement à la construction d'une théorie intéressante. Il y a une certaine forme et organisation dans la conscience à un instant, c'est-à-dire entre les « objets » de la conscience à cet instant, qu'il importe de formuler dans les données dont part la théorie. Pour cela, il faut se munir d'une notion de *structure* qui soit aussi *neutre* que possible. Le paragraphe suivant en présente une.

2.1.3 Notions pour l'analyse de l'état instantané de la conscience : les notions de structure, d'objet et de relation

Où l'on introduit les notions techniques de structure, d'objet et de relation, une structure consistant en des objets et en des relations entre eux. Ces notions permettent une représentation neutre de l'état de la conscience à un instant particulier comme une structure comportant, comme objets, tout élément, aspect, entité ou chose qui est dans l'état de la conscience à cet instant, et comme relations, tous les rapports entre ces éléments, aspects ou choses en tant qu'ils figurent dans cet état de la conscience. On appelle cette structure la structure instantanée de la conscience.

L'état de la conscience à un instant est dans une certaine mesure organisé. Il a pour ainsi dire une « structure ». Les données que l'on emploie pour comprendre les aspects mentaux ayant trait à la fiction devraient comporter quelque chose de cette « structure ». Or, il faut une manière de formuler ces aspects « de structure » ou « d'organisation » de l'état instantané de la conscience en respectant les deux maximes énoncées ci-dessus, c'est-à-dire, il faut une formulation qui soit assez précise pour être utile (*fécondité*), mais qui ne trahisse pas les données phénoménologiques par l'ajout de suppositions « externes » excessives (*fidélité*). Les trois notions – d'abord techniques – de *structure*, d'*objet* et de *relation* fourniront un vocabulaire convenable à l'expression ou la représentation féconde et fidèle de l'état instantané de la conscience.

Les notions de structure, d'objet et de relation se définissent les unes par rapport aux autres de la manière intuitive suivante :

*Une structure consiste en des objets et en des relations entre ces objets*¹².

Il faut tout de suite préciser que le terme « relation » est employé ici pour parler d'un seul rapport qui relie un certain nombre d'objets dans la structure, plutôt qu'une collection de rapports du même « type », chacun reliant une suite ou une collection différente d'objets. Par exemple, dans la structure d'une généalogie (qui a pour objets des individus), le fait pour un individu d'avoir deux frères consiste en deux relations : l'une entre l'individu et l'un de ses frères, l'autre avec l'autre frère. Ce qui n'empêche que des relations différentes pourraient être du même *type* et qu'une structure pourrait comporter plusieurs *types de relations*. Par exemple, dans la structure de la généalogie, « frère de » est un type de relation, « mère de » en est un autre.

Par ailleurs, cette définition n'affirme pas que les objets (ni les relations) forment un ensemble : en fait, comme on le verra, les structures correspondant aux états instantanés de la conscience sont des structures *vagues*.

En outre, cette définition ne prétend pas non plus définir la notion de structure *à partir de* celles d'objet et de relation, ni les notions d'objet et de relation *à partir de* celle de structure : elle est une définition des *trois notions à la fois*. C'est-à-dire qu'il s'agit d'une définition *simultanée* des trois notions *par leurs rapports les unes aux autres*, à la façon de la définition hilbertienne des concepts de point, de droite et de plan, l'un en termes des autres, laquelle consiste à écrire des énoncés où figurent les trois à la fois (ou au moins deux des trois dans chacun) et à poser qu'un point est ce qui se rapporte de cette manière à d'autres choses, qui sont les droites et les plans¹³. Ces notions sont définies ensembles, elles sont *co-définies* pour ainsi dire, de telle sorte qu'il n'y a aucune possibilité de creuser un écart entre elles. Notamment, on frôle la confusion lorsque l'on dit qu'il y a des objets qui manquent à une certaine structure : les objets dont on parle alors sont des objets *d'une autre structure* (« plus grande »). Pareillement si l'on affirme que la structure comporte plus que ses objets et ses relations : les objets et les relations constituent par définition une structure, éventuellement différente de celle dont on parle. Selon cette définition donc, *la structure n'est autre que ses objets et ses relations*, et inversement, *les objets et les relations (pris ensemble) ne sont autres que la structure qu'ils constituent*.

¹²L'arité de la relation – le nombre d'objets qu'elle relie – peut changer entre relations. C'est une question de détail plutôt que de fond, laquelle est dans une grande partie conventionnelle. On permet, pour la commodité de l'exposé, des relations qui prennent un seul objet, ce que l'on appelle souvent propriétés ou prédicats.

¹³Hilbert (1899, §1), et Frege and Hilbert (1900, p226) : « je vois dans mon explication de §1 [de Hilbert (1899)] la définition des concepts de point, de droite et de plan, si l'on prend comme caractéristiques tous les axiomes des groupes I-V. »

Si l'on s'attarde sur ce point – et sur cette définition d'apparence banale – c'est pour distinguer l'acception que prennent les termes « structure », « objet » et « relation » dans ce travail des sens et des connotations qui leur sont associées ailleurs dans la littérature. Ces connotations introduisent des suppositions théoriques implicites, transgressant alors la maxime de *fidélité*, alors qu'ici il n'est question que d'un usage *technique* et donc *neutre* de ces termes. Souvent, en employant les termes « structure », « objet » et « relation » pour parler des systèmes, au sens le plus large, on énonce des thèses telles : ceci est *la structure* du système, *donc* les choses et les aspects de ce système sont exactement les objets et les relations *qui figurent dans cette structure*, ou : ceux sont les objets et les relations du système, *donc* la structure de ce système est *la structure qui consiste exactement en ces objets avec ces relations entre eux*. Par exemple, dans la théorie de la signification, il y a une thèse (la thèse dite « holiste ») qui affirme que la signification d'un terme (objet) n'est autre que sa position dans la toile des rapports entre les termes (structure), et une thèse opposée (la thèse dite « atomiste ») qui affirme que la signification d'un terme (objet) est donnée indépendamment de sa position dans cette toile (structure)¹⁴. On résume parfois cette opposition avec la devise suivante : le holiste affirme la primauté de la structure (la toile) sur l'objet (la signification d'un terme) alors que l'atomiste affirme l'indépendance (sinon la primauté) de l'objet (la signification) par rapport à la structure (la toile). Or, si les termes « structure » et « objet » sont employés dans les sens précisés ci-dessus, et d'ailleurs en vigueur dans cette thèse, ces formules sont du non sens¹⁵. Il n'y a pas de primauté de structure par rapport à l'objet, ni de primauté ou d'indépendance de l'objet par rapport à la structure. La structure et l'objet et la relation sont sur un pied d'égalité. Leur usage dans ce travail ne dissimule ni présuppose aucune *thèse* (holiste, atomiste ou autre).

Ces notions de structure, d'objet et de relation – tant *formelles* pour être manipulées et employées, tant *neutres* pour être fidèles – permettent d'exprimer les données de la présente approche, à savoir les états instantanés de la conscience.

Pour un état instantané de la conscience, on dira que la *structure* de cet état est la structure qui a pour *objets* toute entité, chose, aspect quelconque (réel ou imaginé, « existant » ou « non existant » *etc.*) qui apparaît d'une manière quel-

¹⁴Pour une brève introduction à ce débat, voir par exemple Block (1999).

¹⁵En vérité, le holiste affirme que la structure reliant des termes est égale à celle qui relie les significations des termes, alors que l'atomiste le nie : selon lui, la structure reliant les significations des termes ne comporte aucune des relations de la structure reliant les termes eux-mêmes.

conque dans cet état de la conscience, et pour *relations* les rapports quelconques entre ces objets en tant qu'ils apparaissent dans cet état de la conscience. Pour abrégé la discussion, on emploiera souvent l'expression « structure instantanée de la conscience » au lieu de « structure de l'état instantané de la conscience » ; de manière pareille, on emploiera souvent l'expression *une* structure instantanée de la conscience comme abrégé de « la structure de l'état instantané de la conscience à un instant » (donc, différentes structures instantanées de la conscience ont trait à des instants différents).

Les données dont part la théorie que l'on propose ici sont les structures instantanées de la conscience du sujet. En conséquence de la neutralité et la généralité (et corrélativement, la faiblesse) des notions de structure, d'objet et de relation, *tout* ce qui apparaît dans l'état instantané de la conscience, *tout* ce qui fait partie de son unité plus ou moins organisée, appartient à la structure, tantôt comme objet, tantôt comme relation. Donc, s'agissant de la conscience du lecteur à un instant donné de sa lecture, la structure comporte quelques mots du texte dans son champ visuel et les relations spatiales entre eux (par exemple), aussi bien que quelques personnages, situations, événements du livre (en image, en langage ou autrement) et certaines relations entres eux, aussi bien éventuellement que des termes (apparaissant dans le texte et autres) en tant qu'entités linguistiques, et les relations entre eux ... En un mot : si un élément apparaît dans la conscience du lecteur (de quelque manière que ce soit) à un instant, alors il est dans la structure instantanée de la conscience relative à cet instant.

Il s'ensuit naturellement que la notion d'*objet* prend un sens aussi large que possible ; sens qui pourrait d'ailleurs être rapproché des sens vastes du *Gegenstand* de Meinong, de l'*Objekt* de Husserl ou du *term* de Russell¹⁶. Néanmoins, le fait d'entendre la notion d'objet dans cette acception très large n'autorise pas à l'employer de façon arbitraire, car les objets en question à un moment particulier sont ceux qui *figurent dans la structure en considération* (et eux seulement). En particulier, ce serait une confusion de dire que, puisque la notion d'objet a un sens très vaste, les relations d'une structure sont également des objets. On parle là plutôt d'une *autre* structure, comportant, par-delà les objets et les relations de la structure originelle, d'*autres* objets qui « correspondent » en quelque sens aux relations de la structure originelle. Bien entendu, le sens dans lequel les nouveaux objets « sont » les relations de la structure originelle est difficile, étant plus *intuitif* que *formel* ; c'est d'ailleurs pour cette raison que l'on choisit de parler de « correspondance » plutôt que d'« identité » (on revient

¹⁶Meinong (1904), Husserl (1913a, §37 (notamment, p66, note 1)), Russell (1937, §47).

sur ce point dans §§2.2.1 et 2.4). Il reste que, puisque la distinction entre ce qui compte comme relation et ce qui compte comme objet est toujours relative à la structure qu'ils constituent ensemble, le fait de parler des relations comme si elles étaient des objets implique un glissement vers une autre structure (dont on parle effectivement des objets).

L'état instantané de la conscience est donc représenté ou conçu comme une structure composée d'objets et de relations, dans le sens technique précisé ci-dessus. Dans la section suivante, il s'agit d'approfondir la compréhension de la structure de l'état instantané de la conscience en en remarquant quelques caractéristiques immédiates mais importantes.

2.2 Aspects de l'état instantané de la conscience

Il y a *en gros* trois points relatifs à la structure instantanée de la conscience qui s'avèrent cruciaux pour comprendre la conscience, et d'ailleurs la psychologie de la fiction à partir d'elle. Chaque point consiste en une caractéristique de la structure instantanée de la conscience, laquelle se rapporte aux intuitions fondamentales à l'égard de la conscience. Premièrement, il s'agit d'accentuer la *totalité* de chaque structure relative à elle-même (tout *autre* objet ou relation fait partie d'une *autre* structure), laquelle sera mise en rapport avec la *localité* de l'état instantané conscience (l'absence éventuelle d'*autres* objets « présents » à d'*autres* moments dans l'état instantané de la conscience). Deuxièmement, la distinction entre objet et relation au sein de la structure instantanée de la conscience sera mise en rapport avec celle, intuitive, entre les aspects *au front* et *au fond* de la conscience, ou entre les aspects *au point* ou en *focus* et ceux qui ne le sont pas. Troisièmement, on constate le vague de la conscience, discute sa conséquence pour la structure instantanée de la conscience, et le met en rapport avec la dynamique de la conscience. Les deux premiers points seront abordés dans cette section. La troisième, qui débouche sur une discussion de la dynamique des états instantanés de la conscience, appartient à la section suivante.

2.2.1 La localité de la structure instantanée de la conscience

Où l'on met en évidence la localité de l'état de la conscience : le fait qu'il y a des faits, des images, des aspects, des événements qui ne sont pas dans l'état de la conscience à un instant alors qu'ils y sont présent à un autre instant. Il est cependant une question complexe

de formuler correctement et rigoureusement cette propriété, car la totalité de la structure instantanée par rapport à elle-même – le fait qu'elle n'est autre que ses objets et ses relations – empêche de parler des objets ou des relations qui n'appartiennent pas à cette structure. Cette totalité souligne le besoin d'une notion de correspondance entre des objets et des relations des structures instantanées différentes, et la difficulté de cette notion. Une fois supposée une correspondance entre structures, leur localité peut être exprimée comme le fait d'y avoir, entre deux structures instantanées à deux instants différents, des objets de l'une qui ne correspondent à aucun objet de l'autre.

Une première caractéristique évidente de l'état instantané de la conscience à un instant a trait à son rapport aux états instantanés de la conscience à d'autres instants. *Intuitivement*, on sait très bien qu'il pourrait y avoir faits, images, phénomènes, événements et ainsi de suite, qui sont *présents dans* l'état instantané de la conscience à un moment, mais *non pas dans* l'état instantané de la conscience à un autre moment¹⁷. À un moment pendant la lecture, il est probable que le fait qu'un vieil ami a des yeux bleus ne soit pas présent dans ou n'appartienne pas à l'état instantané de la conscience, alors qu'il appartient à l'état instantané de la conscience à un autre moment. Peut-être le fait que les personnages fictionnels n'existent pas, ou qu'une affirmation du texte est réellement fausse, ne sont-ils pas présents à ou n'appartiennent-ils pas à l'état instantané de la conscience à tel instant pendant la lecture, alors qu'ils appartiennent à l'état de la conscience à d'autres moments.

Ce fait, si banal soit-il, est d'importance capitale. On dit que *l'état instantané de la conscience est local*. Car, par rapport à une sorte de « collection » de tous les faits, aspects, événements et ainsi de suite qui appartiennent à l'état instantané de la conscience à un instant quelconque, les objets dans l'état instantané à un instant particulier n'occupent qu'une petite partie¹⁸.

Ce qui est *intuitivement* vrai pour les faits, événements, images, aspects et phénomènes dans les *états instantanés de la conscience* s'applique également aux *objets* et aux *relations* dans les *structures* de ces états, à quelques réserves près. Certes, intuitivement, pour une paire d'instant distincts, il y a souvent (sinon toujours) des objets de la structure de l'état instantané de la conscience à un des instants qui ne sont pas des objets de la structure de l'état instantané de

¹⁷Le sens « internaliste » de ce « dans » est précisé dans §2.1.1.

¹⁸Naturellement, la mise en jeu d'une telle « collection », et l'emploi du terme local, est *heuristique* et *intuitive* plutôt que sérieuse : on ne prend pas partie sur la question de la possibilité ou la fécondité de penser en termes d'une telle « collection ».

la conscience à l'autre instant ; *idem* pour les relations. En ceci, il y a *localité* : relative à une sorte de « collection » de tous les objets de toutes les structures de l'état instantané de la conscience à tous les instants, les objets de la structure instantanée de la conscience à un instant particulier ne recouvrent qu'une petite partie¹⁹.

Toutefois, bien que intuitivement exact, il est plus complexe d'*exprimer correctement* cette localité des structures instantanées de la conscience. Car, selon la définition de structure (§2.1.3), la structure *n'est autre que* ses objets et ses relations, de sorte que, comme on l'a déjà souligné, il est *stricto sensu* un non sens de dire qu'il y a un objet qui n'appartient pas à la structure. *La structure, et a fortiori la structure de l'état instantané de la conscience, est une totalité par rapport à elle-même*. Néanmoins, comme on l'a expliqué plus haut, on peut comprendre un énoncé au sujet d'autres objets comme un énoncé au sujet d'une *autre* structure comportant non seulement les objets et les relations de la structure originelle mais des objets supplémentaires. Ces deux structures sont différentes, mais on conçoit naturellement une certaine « relation » entre elles qui met les objets de la structure originelle en relation avec les objets de l'*autre* structure qui y *correspondent* (et pareillement pour les relations). Lorsque l'on dit qu'un objet n'appartient pas à la structure originelle, on veut dire (*précisément*) qu'un objet de cette autre structure n'est mis en relation avec aucun objet de la structure originelle par cette « relation naturelle » entre elles. Ce faisant, on se met dans une perspective plus grande – relative à laquelle la structure originelle n'est plus une totalité – où l'on aperçoit *intuitivement* les objets et les relations des deux structures et les « correspondances *intuitives* » entre eux²⁰.

La *totalité* de la structure, et notamment de la structure de la conscience, par rapport à elle-même a donc pour effet de souligner la difficulté de mettre en relation des structures différentes, c'est-à-dire la difficulté de la « *correspondance* », ou comme on dit souvent, l'« *identification* »²¹ entre des structures différentes, et notamment entre les structures de la conscience à des instants différents. En effet, la question de la « mise en correspondance » ou de « l'identification » des

¹⁹Rappelons de §2.1.3 que « la structure instantanée de la conscience » est un abrégé pour « la structure de l'état instantané de la conscience ».

²⁰Techniquement, on a affaire à une structure « élargie », dont les objets et les relations « *correspondent intuitivement* » aux objets et aux relations des deux structures originelles, où il y a des relations entre ces objets qui « identifient » certains objets de l'une des structures originelles avec certains objets de l'autre, et pareillement, *grosso modo*, pour les relations.

²¹Comme dans le cas de « correspondance » entre objets et relations d'une structure, discuté dans §2.1.3, on emploie souvent, par souci de limpidité, le terme « correspondance » plutôt qu'« identité ».

objets (percepts, phénomènes, faits, événements *etc.*) appartenant aux structures instantanées de la conscience à des instants différents est particulièrement épineuse, surtout lorsque les instants en jeu sont éloignés dans le temps. Pour l'enfant qui voit *Notre Dame* pour la première fois, il y a dans l'état instantané de sa conscience un *objet* qu'on « identifierait » à *Notre Dame*, pourtant il y aura également un objet qu'on appellerait *Notre Dame* dans l'état de la conscience de la même personne en adulte, revoyant *Notre Dame* des années plus tard : en quel sens sont-ils *les mêmes* ces deux objets, que l'on « identifie » ordinairement sous le nom de *Notre Dame* ? En quel sens les événements de *Le Bruit et la fureur* sont-ils *les mêmes* dans la conscience du lecteur entre la première et la seconde lecture ?²² Cependant, la *localité* de la structure de la conscience ne peut être exprimée qu'après avoir adopté une telle *correspondance* entre les structures de l'état instantané de la conscience à des instants différents : il s'agit du fait que, entre deux structures à deux instants différents, il y a des objets de l'une qui ne *correspondent* à aucun objet de l'autre.

On reviendra sur la difficulté en « mettant en correspondance » ou en « identifiant » les objets, et d'ailleurs les relations, des structures instantanées de la conscience dans §2.4. Pour le moment, il convient de parler des objets des structures instantanées de la conscience, et donc des états instantanés de la conscience, avec les termes « ordinaires », en supposant qu'on les a déjà mis en correspondance, aussi bien entre les structures différentes qu'avec nos propres notions en tant que théoriciens et observateurs. Donc, on parle de l'objet *Notre Dame* dans la structure instantanée de la conscience à l'instant où le sujet regarde *Notre Dame*, en supposant l'identification de cet objet « comme » *Notre Dame* et l'identification de cet objet avec d'autres objets *Notre Dame* dans d'autres structures instantanées de la conscience.

2.2.2 Objets et relations : le focalisé et le non focalisé

Où l'on reconnaît la différence entre les objets et les relations d'une structure instantanée de la conscience comme celle entre les aspects qui sont focalisés dans la conscience à ce moment et ceux qui, quoique dans la conscience, y sont non focalisés.

À un moment, le sujet regarde sa bibliothèque, comptant ses livres. Dans la structure de la conscience à ce moment, il y a ce livre-ci, ce livre-là, et ainsi

²²De Faulkner, ou, autre exemple, *Un drame bien parisien* d'Allais, dont Eco affirme qu'il « a été écrit pour être lu deux fois (au moins) : la première lecture présuppose un Lecteur Naïf, la seconde un Lecteur Critique qui interprète l'échec de l'entreprise du premier. » (Eco, 1985, p255)

de suite ; ils sont des objets de la structure instantanée de la conscience à ce moment. Ce sont ces livres qui, dans un certain sens, « intéressent » le sujet. Pour le reste, il y a des relations entre les livres (objets), telles les relations spatiales – ce livre est à côté de cet autre. Dans un certain sens, ces rapports *n'intéressent pas* le sujet. Toutefois, elles sont importantes pour ce qui l'intéresse, car il ne pourrait pas compter ses livres s'il ne pouvait pas distinguer entre eux grâce à ces relations spatiales. À un moment (pour prendre un autre exemple), le sujet a affaire à une rencontre entre deux individus qui se saluent de façon conviviale et familière, pour commencer à discuter leurs familles ou les derniers événements de leurs vies. Les deux individus sont présents comme objets dans les structures instantanées de la conscience à ces moments : ils sont « en avant » ou « dans le *focus* » de « l'attention » ou de « l'intérêt » du sujet. Pour le reste, il y a dans ces structures instantanées de la conscience à ces moments une relation (la familiarité, la connaissance) entre ces deux individus, relation qui n'est pas, quant à elle, « en avant » dans « l'attention » ou dans « l'intérêt » du sujet. Néanmoins, comme dans l'exemple précédent, cette relation est importante dans la mesure où elle permet (ou explique, ou laisse comprendre) les comportements des individus.

La conclusion que l'on tire de ces exemples concerne le *statut* des objets et le *statut* des relations d'une structure instantanée de la conscience. Les objets d'une structure instantanée de la conscience sont les aspects qui sont « en avant » ou « au point » dans l'état instantané de la conscience à cet instant ; on dit qu'ils sont des aspects *focalisés* dans cet état instantané. En revanche, les relations sont les aspects qui sont, quoique éventuellement importants, « au fond » ou pas « au point » dans l'état instantané à cet instant ; on dit que ces aspects sont *non focalisés*.

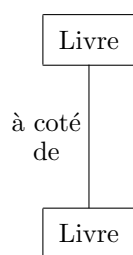
On comprend mieux cette différence de statut en comparant les structures décrites ci-dessus avec des structures instantanées où les aspects *non focalisés* auparavant sont maintenant *focalisés*, où, pour ainsi dire, le sujet s'y intéresse. Il s'agit d'un moment où l'admirateur de sa propre bibliothèque en vient à porter attention au fait que deux livres sont à côté l'un de l'autre. La structure instantanée de la conscience à ce moment comporte toujours les deux livres, en tant qu'objets, et la relation « être à côté de » entre eux. Mais quelle est la *différence* entre cette structure instantanée de la conscience et la précédente, lorsque le sujet ne portait pas attention à la situation spatiale respective des deux livres ? C'est que cette relation spatiale est venue prendre une « place distincte » dans la structure instantanée de la conscience, analogue à la « place

distincte » d'un des livres lorsqu'il pense précisément à ce livre. D'après les définitions de §2.1.3, on a un terme précis pour cette « place distincte » : il y a dans la structure instantanée à cet instant non seulement la relation d'être à côté l'un de l'autre mais également un *objet* qui y « *correspond* ». Rappelons, de §2.1.3, qu'alors que l'on est tenté de dire que cet objet « est » cette relation, le sens de cette « identité » pourrait s'avérer difficile et imprécis, si bien que l'expression plus prudente de « correspondance » paraît plus convenable. Pour distinguer entre la relation et l'objet, on utilisera les guillemets pour la première (« être à côté de ») et le nom, en italiques, pour le dernier (*l'adjacence*). L'objet (*l'adjacence*) qui « correspond » à la relation « être à côté de » est relié aux deux autres objets (les *livres*) par une relation comprise intuitivement comme « entrent dans la relation » : les deux livres entrent dans la relation d'adjacence.

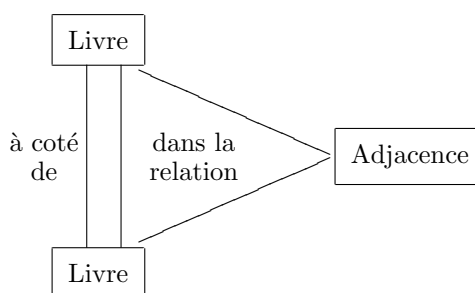
Formellement : soit $A(,)$ la relation « ... est à côté de ... », l_1 un des *livres* (objets), l_2 l'autre. Dans la structure instantanée de la conscience du premier exemple – où le caractère adjacent des livres n'est pas *focalisé* – la relation A s'applique aux objets l_1 et l_2 : $A(l_1, l_2)$. Dans le deuxième exemple, où l'attention est prêtée à la relation entre les objets, il y a un troisième objet, a (*l'adjacence*), qui « *correspond* » à la relation « ... est à côté de ... », et une autre relation R , « ... et ... entrent dans la relation ... », telle que $R(l_1, l_2, a)$.

En diagramme :

Premier exemple



Deuxième exemple



L'*esprit* de la distinction en jeu ici est familier. Il convient de l'éclairer en s'appuyant sur des analogies avec d'autres distinctions héritées de la tradition. Même si chacune de ces analogies est imparfaite, il ne s'agit pas de la discuter en détail pour cerner les différences et les ressemblances, car, en fin de compte, le but est de faire allusion à *l'idée générale* dont il s'agit ici, qui est commune, dans une manière ou dans une autre, à toutes les distinctions discutées ci-dessous.

Prenons d'abord une analogie avec la logique, laquelle met en rapport d'un côté les objets et les noms²³, et de l'autre côté les relations et les prédicats²⁴,²⁵. Selon cette analogie, les livres l_1 et l_2 correspondent aux noms, la relation A correspond à un prédicat ; le langage muni de ces noms et ce prédicat correspond au premier exemple (ou le sujet ne prête pas l'attention à la relation « à côté de »). On peut *dire* dans ce langage que les deux livres sont à côté l'un de l'autre – $A(l_1, l_2)$ – mais on *ne peut pas parler* (dans ce langage) directement de la relation « être à côté » (A) – on ne peut pas dire par exemple que c'est *cette* relation qui subsiste entre ces deux objets l_1 et l_2 . Car, alors qu'un nom pourrait « tomber sous » un prédicat (l_1 et l_2 sous A), un prédicat ne peut pas « tomber sous » un autre prédicat, à moins que ce soit un prédicat dit « d'ordre supérieur »²⁶. Et on manque aussi bien un prédicat d'ordre supérieur qu'un nom pour le prédicat A (« à côté de »). Le second exemple – où il y a un objet a (l'*adjacence*) qui « correspond » à la relation « à côté de » – diffère exactement dans ce sens. Selon l'analogie, il s'agit d'un langage étendu, muni d'un *nom* pour la relation « à côté de », auquel correspond l'objet a (l'*adjacence*), et un *prédicat* auquel correspond la relation R (« entrent dans la relation »). Dans ce langage, on peut dire que les deux livres entrent dans la relation d'adjacence : $R(l_1, l_2, a)$. Plus généralement, on peut *parler explicitement* de la relation « à côté de » puisqu'elle a un nom dans le langage. Avec un nom, on peut « parler explicitement » de ce à quoi il réfère ; de la même façon, lorsqu'il y a un objet dans la structure instantanée de la conscience, le sujet en « pense explicitement ». En revanche, avec un prédicat, on peut « l'employer » sans « parler explicitement » de la relation ou de la propriété à laquelle il réfère ; de même, une relation dans une structure instantanée est « présente » ou « employée » ou « effective » sans que l'on « en pense explicitement » ou que l'on « y prête explicitement l'attention ».

On pourrait exprimer ce même point dans le langage suggestif du *Tractatus*. Dans son optique, « la proposition élémentaire [est une] fonction de noms sous la

²³On suit la suggestion de Quine (1960, p180), selon lequel « les termes singuliers, autres que les variables, qui sont traités comme simples dans ce sens peuvent de façon suggestive être appelés des *noms*. » Cette terminologie est certainement plus « suggestive » que la terminologie formelle : « constantes ».

²⁴Afin d'écartier des confusions terminologiques possibles, on emploie le terme « prédicat » plutôt que « relation » pour parler de l'équivalent logique des relations sous cette analogie.

²⁵Pour parler en termes fregéens, il s'agit d'une analogie entre les objets et les objets fregéens, et entre les relations et les concepts fregéens (si l'on met à côté la question des concepts d'ordre supérieur) ; voir Frege (1892).

²⁶Comme le dit Frege, *op cit.*, p201 : « Les concepts du second ordre, qui s'appliquent aux concepts, diffèrent de manière essentielle des concepts du premier ordre, qui s'appliquent aux objets. »

forme : ' fx ', ' $\phi(x, y)$ ', *etc.* »²⁷, d'où, on pourrait employer la même analogie en rapprochant les objets (l_1, l_2, a) des noms (wittgensteiniens) et les relations (A, R) des fonctions (wittgensteiniennes). Or, avec des noms, on peut représenter ou nommer des objets, on peut « parler d'eux »²⁸. Il ne semble pas (au moins dans le cadre du *Tractatus*) que les fonctions réfèrent, ou que l'on puisse « parler de » ce à quoi elles réfèreraient. Au mieux, on peut dire, avec la proposition qui est la fonction appliquée aux noms appropriés, comment les choses sont (*wie ein Ding ist*) relativement aux objets auxquels réfèrent ces noms²⁹. Par exemple, avec la proposition $A(l_1, l_2)$, on dit comment les choses sont relativement aux deux livres (ils sont à côté l'un de l'autre). Donc, sans un objet, ou, conformément à l'analogie en vigueur, sans un nom qui « correspond à » ou « nomme » la relation « à côté de », on ne peut pas *dire* des choses au sujet de cette relation, on ne peut pas *parler de* elle. Dans ce sens, la relation est « inexprimable ». Or, selon Wittgenstein, « il y a assurément de l'inexprimable. Celui-ci se *montre*. »³⁰ Ainsi, si l'on ne peut pas dire des choses au sujet de la relation, on peut néanmoins *montrer* des choses à son égard, ou, pour déformer un peu le langage, on peut *montrer de* elle. Bref, de l'objet, on rapproche le *dit* ; de la relation, qui est *là* mais ne peut pas être *dite*, on rapproche le *montré*.

Voici quelques autres distinctions auxquelles une référence pourrait faciliter la compréhension de la teneur de la distinction entre objet et relation. La distinction que tire Dennett entre l'*explicite* et le *tacite*, inspirée de la distinction ryléenne entre *savoir* (*know that*) et le *savoir-faire* (*know how*) : « le savoir-faire [que Dennett appelle *tacite*] doit être intégré dans le système dans une manière qui ne requiert pas qu'il soit représenté (de manière *explicite*) dans le système »³¹. L'objet dans la structure instantanée de la conscience est *explicite*, la relation (sans objet qui y « correspond ») est « intégrée » dans la structure sans être objet : elle se rapproche du *tacite*. Apparenté à l'idée de la distinction entre explicite et non explicite, il pourrait convenir de mentionner l'opposition entre manifeste et non manifeste : l'objet est manifeste dans la structure instantanée de la conscience, alors que la relation, bien *qu'active*, ne l'est pas ; de plus, pour la rendre manifeste, il faut recourir à une autre structure avec un objet (manifeste) qui lui « correspond ».

Pour terminer, il est besoin de faire quelques remarques au sujet d'un rapprochement séduisant avec la distinction que souligne Husserl entre le « ce vers

²⁷Wittgenstein (1922, §4.24).

²⁸*Op cit.*, §3.221.

²⁹*Ibid.*

³⁰*Op cit.*, §6.522.

³¹Dennett (1987, p218) (accent ajouté). La référence à Ryle est (2000, Ch. II). On remarque par ailleurs que certains auteurs emploient le terme « implicite » à la place de « tacite ».

lequel la conscience est tournée », lequel est « perçu « attentivement » ou noté accessoirement », et le « champ d'intuition » qui forme « l'arrière-plan »³². Dans le cas de la structure instantanée de la conscience qui comporte un objet qui « correspond » à une relation « à côté de », il semblerait que le sujet soit « tourné vers » ou soit « attentif au » fait que les livres sont adjacents. Pourtant, il faut se méfier d'un rabattement précipité et rude du couple objet / relation sur le couple premier plan / arrière-plan. Comme on verra dans §2.3.1, la collection d'objets d'une structure instantanée de la conscience ne peut pas être identifiée à la collection d'éléments du « premier plan » de la conscience à cet instant, et la collection de relations ne peut pas être identifiée à la collection d'éléments de l'arrière-plan. Le rapport entre les deux couples est plus complexe.

Du reste, il serait imprudent d'identifier l'opposition entre objet et relation à n'importe quelle opposition empruntée à la tradition. Car, tout simplement, d'autres théories s'enfoncent dans leurs intérêts, leurs problématiques, leurs présupposés et leurs préjugés propres, de sorte que tout rapprochement rigoureux exige un grand nombre de réserves spécifiques, et un grand travail pour en faire la liste. En outre, pour autant que les cadres philosophiques évoqués puissent s'appliquer à l'état instantané de la conscience, ils risquent de transgresser la maxime de fidélité qui motive la conception « neutre » de structure, objet et relation. Husserl par exemple prend position par rapport à la « méréologie de la conscience », c'est-à-dire par rapport à la constitution et la composition des « actes » de la conscience et par conséquent sur les objets dans la conscience (au sens de « dans » précisé dans §2.1.1)³³. Pour prendre un autre exemple, on pourrait lire dans des propos de Wittgenstein au sujet du langage, interprétées comme s'appliquant à la conscience, des affirmations que la structure instantanée de la conscience a une certaine structure linguistique voire logique³⁴. Or, une structure méréologique n'est autre qu'une *structure*, dans le sens faible explicité dans §2.1.3, qui *satisfait à certaines conditions supplémentaires*, et une structure logique n'est autre qu'une *structure*, au sens faible employé ici, qui *satisfait à d'autres conditions supplémentaires*³⁵. Dans la mesure où elles posent des structures *plus fortes* que celles qui sont employées dans ce travail, les propositions telles celles de Husserl ou de Wittgenstein doivent être traitées comme des *thèses* (contrairement à leurs intentions). Il se pourrait bien que ces thèses

³²Husserl (1913a, p62).

³³Husserl (1913b, IV^e et V^e Recherches).

³⁴*Op cit.*, §5.6 : « Les limites de mon langage signifient les limites de mon propre monde. »

³⁵Pour quelques conditions méréologiques, voir Husserl, *op cit.* IV^e Recherche, ou Simons (1987). Pour les conditions logiques, voir Wittgenstein *op cit.* ou n'importe quel manuel de logique.

soient vraies, pourtant il ne reste pas moins qu'elles sont des thèses. Puisqu'il s'agit ici *seulement* d'établir une manière de représenter l'état instantané de la conscience qui satisfait autant que possible à la maxime de *fidélité* (§2.1.2), on ne pourrait pas admettre ces thèses à ce stade de l'enquête sur la conscience. La *neutralité* du cadre de ce travail, des notions de structure, d'objet et de relation, et par conséquent la *neutralité* de la distinction qu'il s'agit de tirer ici entre objet et relation : voilà la différence la plus significative et la plus cruciale entre la distinction visée ici est celles qu'on a évoquées ci-dessus.

2.2.3 Résumé : le triplet focalisé / non focalisé / absent

Où l'on résume le propos des deux paragraphes précédents dans le triplet de statuts qu'un aspect pourrait avoir relativement à un état instantané de la conscience. Il pourrait être tantôt présent et focalisé, tantôt présent et non focalisé, tantôt absent. À la différence de certaines distinctions que l'on trouve dans la tradition, ce triplet ne représente pas une gradation linéaire de modes de présence dans la conscience.

Sous les réserves relatives à la difficulté d'identifier les objets (et d'ailleurs les relations) entre différentes structures instantanées de la conscience, soulignées dans §2.2.1, aussi bien que d'autres complexités de l'expression mises en évidence dans §2.1.3, on pourrait résumer les conclusions de la présente section dans une expression relâchée mais néanmoins suggestive, de manière suivante. Il y a trois manières pour « quelque chose » d'être « dans » l'état instantané de la conscience *à un instant*. Premièrement, il peut ne pas y être du tout : il peut être *absent* de la conscience à cet instant. C'est une conséquence de la localité mise en évidence dans §2.2.1. Deuxièmement, il peut être présent d'une manière « explicite » ou « remarquée » ou « en relief » : il pourrait être *focalisé*. Troisièmement, il peut être dans l'état instantané de la conscience à ce moment, mais « au fond » ou de manière « non remarquée » ou « non explicite » : il peut être *non focalisé*. C'est la conclusion de §2.2.2 : les objets de la structure instantanée de la conscience à cet instant sont focalisés, les relations sont non focalisées.

Cette distinction rappelle une triple structure de présence dans la conscience que l'on trouve, si ce n'est qu'implicitement, chez les psychologues ou les philosophes qui ont traités de l'attention ou de la conscience. On a déjà remarqué que James et Husserl reconnaissent qu'aussi bien que les aspects « au centre d'attention » ou « explicites » ou « vers lesquels la conscience est tournée »,

il y en a d'autres qui sont « en marge » ou « à l'arrière-plan »³⁶. Mais, à un instant donné, ils admettent également des aspects ou des objets qui étaient « dans » l'état instantané de la conscience à un autre instant, mais qui n'y sont pas présents à l'instant en question, tels des souvenirs non re-présentés à cet instant. C'est-à-dire, il y a une distinction entre ce qui est absent, ce qui est « au centre de l'attention », et ce qui est en « marge ». Pour reprendre un exemple d'un psychologue, on a déjà mentionné la distinction de Cowan entre le *focus attentionnel* et l'ensemble de « nœuds mentaux » qui ont un niveau d'activation non nul³⁷. Mais il y a un troisième terme : les « nœuds mentaux » qui ont un niveau d'activation nul, qui sont les aspects de la mémoire de l'individu qui ne sont même pas en jeu à l'instant actuel. Le triplet *focus attentionnel*, nœuds d'activation non nulle hors le *focus attentionnel* et nœuds d'activation nulle rappelle certainement le triplet *focalisé, non focalisé, absent*.

Toutefois, il y a une différence capitale entre le triplet proposé ici est ceux que l'on a repérés dans la tradition. À savoir, alors que les triplets de la tradition admettent ou supposent une *gradation* dont les trois termes du triplet sont des moments successifs, *il n'y a pas de telle gradation dans le triplet focalisé, non focalisé, absent*. Chez Husserl ou James, il y a apparemment une « graduation d'attention » ou « d'appartenance à la conscience », venant de l'absence, à travers une présence « en marge » ou « en arrière-plan », à une présence « au centre » ou « explicite » ; chez Cowan, la graduation de niveaux d'activation instaure une gradation qui vient de l'activation nulle, à travers les niveaux d'activation non nuls mais petits, qui sont le plus souvent absents du *focus attentionnel*, à un niveau d'activation significatif, et une présence dans le *focus attentionnel*. Par contraste, comme on verra en détail dans §2.3.1, l'entrée en jeu d'un objet ne passe nécessairement ni naturellement par une relation qui lui « correspond » : le triplet *focalisé, non focalisé, absent* ne permet pas un ordre *linéaire* de passage naturel entre les trois termes.

Néanmoins, le triplet *focalisé, non focalisé, absent* admet une réserve bien connue et rappelée aussi bien par les philosophes (James, Husserl) que par les psychologues (Cowan), à savoir que la distinction entre les termes est *vague*. En effet, on dit souvent que la conscience elle-même comporte un certain caractère vague : il est difficile, à un instant donné, de dire exactement ce qui est « dans » l'état instantané à cet instant et de dire exactement ce qui est *focalisé* et ce qui est *non focalisé*. Un *leitmotiv* de cette thèse est l'importance du vague pour la question de la dynamique : dans la section suivante on discutera aussi bien du

³⁶James (1981, p255), Husserl (1913a, p62).

³⁷Cowan (1993), et la discussion de §1.5.2.

vague de la conscience que de sa dynamique, et du rapport entre eux.

2.3 La dynamique de la conscience

Le problème épineux du rapport entre fiction et réalité a été reconnu être, pour la plupart, le problème du *développement* de ce rapport, de sa *dynamique*. Dans §§1.4 et 1.5, *non seulement* on a mis en évidence l'incapacité des théories traditionnelles de rendre compte du rapport entre l'aspect qu'elles posent comme « fictionnel » et l'aspect « réel » qui l'oppose (*faire comme si* et croyance, entités mentales « fictionnelles » et entités mentales « réelles », processus simulé et processus ordinaire), mais *de plus* ces insuffisances ont indiqué une incapacité de penser la dynamique ou le développement de ce rapport dans les cadres proposés par ces théories. En manquant l'*état instantané* du rapport entre fiction et réalité, elles manquent par là même sa *dynamique*.

Ce point est général et capital : *on ne peut pas concevoir correctement la dynamique sans avoir correctement saisi et représenté l'état instantané dont la succession constitue cette dynamique*. Et, à l'inverse, *on n'a pas conçu proprement l'état instantané si la conception que l'on en emploie ne permet pas une pensée de la dynamique*. Ce constat, d'ordre *méthodologique*, impose des contraintes sur une prétendue théorie psychologique de la fiction. En particulier, une théorie qui décrit le rapport (psychologique) entre fiction et réalité à un instant donné s'assigne un but trop facile si elle ne prend pas également en charge la tâche de la décrire dans une manière qui soit convenable *aussi* pour penser la dynamique de ce rapport. Si d'ailleurs elle se donne cette tâche facile, elle risque d'introduire des éléments qui, quoique « intuitifs », sont nuisibles à une pensée de la dynamique ; ceci semble être le cas des théories considérées dans §§1.4 et 1.5.

Par conséquent, comme on l'a souligné plus haut, si la conscience doit servir de soubassement pour une théorie du rapport entre la fiction et la réalité (dans la psychologie du lecteur), comme elle fait dans le cadre « internaliste » adopté ici, alors il faut que la conception que l'on en emploie prenne suffisamment en compte sa dynamique. En l'occurrence, la représentation utilisée de l'état instantané de la conscience doit convenir à la représentation et à la pensée de la succession de ces états instantanés. Pour répéter en d'autres termes, il faut que l'état instantané *reflète* tant la dynamique pour en permettre la pensée, et que la dynamique *corresponde* tant aux états instantanés pour pouvoir être abordée en termes d'eux³⁸.

³⁸Alors qu'elle est importante, il n'y a pas de place ici pour s'arrêter sur la question du

Le nom pour ce rapport symbiotique entre l'état instantané et la dynamique dont il fait partie est *continuité*. L'état instantané entre dans un rapport « fécond » ou « étroit » avec la dynamique dans la mesure où il comporte en lui-même quelque chose de significatif pour la dynamique. En général, ce quelque chose consiste en un rapport entre l'état instantané de la conscience à un instant et les états instantanés aux instants immédiatement ultérieurs ou antérieurs : il leur *couvre* en partie, il leur *chevauche*, il leur est *proche*. Plus précisément, il y a un *espace*, impliquant des notions de couvrement, de chevauchement, de proximité, tel que les états instantanés à des instants voisins entrent dans des rapports de chevauchement ou de proximité. Le rapport étroit exigé entre l'état instantané et la dynamique consiste *exactement* dans *ces* rapports de chevauchement ou de proximité (relatifs à *cet* espace); c'est ce que l'on appelle *continuité*. Notamment, la notion appropriée de *continuité* est toujours relative à un *espace*, ce qui n'implique nullement que l'espace, ni d'ailleurs la notion de continuité, la représentation de l'état instantané et de dynamique, n'est fixée ou donnée indépendamment des autres notions. Au contraire, il s'agit de formuler une conception unitaire d'un espace, d'une notion de continuité, d'une représentation de l'état instantané et de sa dynamique qui fait l'affaire : qui est telle que les *états instantanés* s'organisent dans l'*espace*, qui implique la *notion de continuité*, relative à laquelle la *dynamique* est continue. La notion de continuité est donc un aspect de cette conception. *L'exigence de continuité*, en revanche, est *l'expression rigoureuse de la contrainte méthodologique sur la représentation de l'état instantané que l'on pourrait utiliser*. Elle requiert que la représentation de l'état instantané va de paire avec un espace et une notion de continuité relativement selon lesquels sa dynamique est continue.

Il s'agit dans cette section de faire quelques pas dans le développement d'une conception qui satisfait à cette exigence, en termes des structures instantanées de la conscience introduites ci-dessus. Bien entendu, étant donné la difficulté du problème de la dynamique, on ne saurait pas attendre une théorie complétée à ce stade; or la discussion devrait mettre en relief la possibilité et la promesse de développer la théorie dans cette direction. A cette fin, l'apport clef sera le caractère vague de la conscience à un instant, qui sera mis au service d'une pensée de la continuité et donc de la dynamique³⁹. On considère premièrement

temps. Pour le propos de cette thèse, on se contente avec une notion « approximative » du temps, avec des « instants » qui ne sont pas infiniment précis, en demandant du lecteur une lecture charitable.

³⁹On n'a malheureusement pas de place pour s'attarder sur l'histoire du rapport et des rapprochements entre le vague et la dynamique, laquelle est apparemment très complexe. Il ne reste pas moins que ce rapport est bien là, comme l'est témoigné par la parenté entre le

le vague de l'état instantané de la conscience et des façons de la représenter (§2.3.1). Dans un deuxième temps (§2.3.2), on s'attarde sur la notion de continuité, sur son sens et sur son importance pour la problématique de la conscience, pour en venir après (§2.3.3) à une approche de la continuité de la succession des états instantanés de la conscience qui s'appuie sur leur vague.

Le rapport *symbiotique* entre état instantané et dynamique atteint dans un travail conceptuel et théorique sous le nom de continuité a le mérite de se débarrasser de certaines « fausses questions », dont celles de savoir si tel ou tel aspect est « synchronique » ou « diachronique ». Car, si l'état instantané est proprement conçu (si, notamment, la dynamique des états instantanés est continue), alors les aspects de cet état à un instant ont une importance quant aux états instantanés à des instants voisins, par la continuité de la dynamique. Autrement dit, les aspects « synchroniques » de l'état instantané ont *également* une portée « diachronique », et vice versa. Pour finir (§2.3.4), on s'arrête sur ce point, aussi bien que d'autres conséquences de la continuité de la dynamique des états instantanés de la conscience, qui s'avéreront utiles dans la suite.

2.3.1 Le caractère vague de la structure

Où l'on remarque le caractère vague de la conscience, et on en rend compte en développant la notion de structure instantanée vague, consistant en une collection ouverte d'objets et une collection ouverte de relations entre eux. Contrairement à un certain nombre d'approches du vague, il ne s'agit pas de ramener le vague au précis (un élément vague à un élément précis, ou à une collection précise d'éléments précis), mais plutôt de le prendre en charge directement. Par ailleurs, le vague de la structure instantanée de la conscience implique une double structure de l'arrière-plan de la conscience à un instant donné : il consiste non seulement en les relations, mais également en des objets dont l'appartenance à la structure instantanée est vague.

La conscience est notoirement éphémère ; il est difficile de dire quoi que ce soit sur son contenu à un instant donné. À un moment de lecture, le lecteur est-il conscient de tel ou tel aspect introduit ailleurs dans le livre ? Cette donnée appartient-elle à son état de la conscience à ce moment ? Et comment : prend-il une place explicite dans l'attention, ou est-il plutôt au fond ? On ne saurait pas

paradoxe de Zénon et celui des sorites, par l'accent depuis Parménide sur l'état instantané (le présent) comme état « clair », et par la rejection de cette position au profit d'une optique dynamique et « flou » par des philosophes tels que Bergson.

toujours donner à ces questions des réponses précises. L'état instantané de la conscience est *vague*. Il s'agit, dans ce paragraphe, de raffiner la notion de structure instantanée proposée ci-dessus pour lui permettre de représenter fidèlement ce caractère vague de la conscience, et de remarquer quelques conséquences immédiates de ce caractère.

Le vague de la conscience est, pour l'essentiel, un vague dans la présence dans l'état de la conscience : la question de savoir si un aspect apparaît dans l'état instantané est parfois vague. Cela vaut aussi bien pour les aspects focalisés (tel personnage apparaît-il dans l'état de la conscience du lecteur à ce moment de sa lecture ?) que pour les aspects non focalisés (tel fait décrit ailleurs dans le livre, qui explique les événements décrits dans le passage lu actuellement, est-il « à l'œuvre » dans la lecture de ce passage, est-il présent, au fond, dans l'état de la conscience à ces moments ?). Ce caractère vague a pour conséquence que, pour représenter ou « exprimer » l'état instantané de la conscience à un instant donné en termes de structures, d'objets et de relations, il n'est pas question d'y associer une structure précise. On ne saurait même pas y repérer une collection précise d'objets ou une collection précise de relations. Donc, pour représenter exactement les états de la conscience, on emploie plutôt des structures *vagues* ou *ouvertes*, qui se définissent comme suit (l'emploi terme « ouvert » sera expliqué prochainement).

*Une structure vague consiste en une collection ouverte d'objets, avec une collection ouverte de relations entre eux*⁴⁰, étant entendu qu'une collection est ouverte s'il est impossible de faire la liste de tous ses membres.

Cette formule n'est pas une définition précise et complète de la notion de vague ou de l'ouvert ; elle figure notamment le mot « impossible », dont le sens n'est pas précisé. Dans cette thèse, cette « impossibilité » sera entendue comme une contrainte *sur le théoricien* de ne pas supposer la possibilité de ramener le vague au précis. Comme telle, elle s'oppose aux « analyses » du vague en termes des entités précises, telle la proposition de Williamson selon laquelle « la proposition exprimée par un énoncé vague dans un cas limite est vraie ou fausse, et nous ne pouvons pas savoir laquelle »⁴¹ ou celle de Fine selon laquelle « un

⁴⁰Il y a deux conséquences du caractère ouvert de la collection des relations qui peuvent être distinguées. D'une part, les types de relations forment une collection ouverte : il n'est pas certain si tel type de relation (par exemple, « à côté de ») relie des objets dans la conscience. D'autre part, pour un type de relation donnée (« à côté de »), qui relie certains objets, les paires (ou n-uplets) des objets reliés forment une collection ouverte. Voir §2.1.3 sur la différence entre la relation et le type de relation.

⁴¹Williamson (1994) et (1997, p921). L'analyse williamsonnienne de notre connaissance

énoncé [comportant des termes vagues] est vrai s'il est vrai dans toutes les façons dont il peut être rendu complètement précis »⁴². Pour exprimer ces positions en termes de collections, Williamson remplace une collection ouverte (ou vague) par une collection précise, alors que Fine identifie une collection ouverte à la collection (*a priori* précise) de « précifications » de cette collection, c'est-à-dire de collections précises que la collection ouverte « pourrait être » (si elle était précise). Dans les deux cas, le vague est ramené au précis : tantôt à une précification particulière (collection précise), tantôt à une collection précise de précifications.

La raison pour ne pas adopter cette réduction du vague au précis tient à deux thèses implicites dans les deux positions : premièrement, ils soutiennent une certaine *possibilité* de trouver le remplacement précis pour le vague (précification ou collection de précifications), deuxièmement, ils supposent ou soutiennent la *fécondité* de cette manière de penser le vague, c'est-à-dire de la démarche qui consiste à le ramener au précis. Quant à la question de la possibilité de trouver une précification appropriée ou une collection appropriée de précifications, on ne s'en occupera ni ne se prononcera là-dessus ici ; en effet, on se permet d'esquiver la question en ne pas *supposant* une telle possibilité. Il est donc un choix stratégique plutôt que théorique – méthodologique plutôt que métaphysique, si l'on peut dire – d'admettre pleinement le caractère vague ou ouvert, de *travailler* avec ce vague comme une *donnée du problème*, plutôt que de la nier ou d'y obvier au profit des formulations précises plus ou moins « possibles ». Cette raison pour ne pas adopter la démarche williamsonienne ou finienne tient donc à la *neutralité* par rapport à la thèse qui soutient une telle possibilité : on met cette thèse *hors sujet* ici. En revanche, leur deuxième thèse – de la fidélité de leur démarche – est carrément *rejetée* ici. En un mot, le rabattement du vague sur le précis *ne convient pas pour penser proprement le vague* (du moins, dans le cas de la structure instantanée de la conscience). On développera cette critique dans §2.3.3, mais ses grandes lignes ont déjà été anticipées : cette approche du vague ne serait pas capable de prendre en compte son rapport à la dynamique.

En prenant en charge le vague en lui-même, on est contraint à parler directement de lui, et des collections ouvertes, ce qui est une entreprise périlleuse, qui frôle souvent le risque de confusion. Il est donc nécessaire, pour éviter quelques

relative à des énoncés vagues, étant formulée dans la logique modale ordinaire, vient ressembler l'analyse finienne du vague lui-même ; pour cette raison, on ignore les questions épistémiques ici.

⁴²Fine (1975, p278). Dans le texte, on expose une version simple du cadre de Fine, dont les subtilités ne sont pas importants ici. Par ailleurs, on reviendra sur ce genre d'approche, et sur la question générale du vague, dans un cadre plus technique dans §5.1.5.

confusions possibles, d'expliciter les significations de certaines expressions employées dans la suite. On dit qu'une collection est *ouverte* puisque la liste de ses membres est *ouverte* : à chaque étape d'une tentative de faire une liste, il y a toujours des (éventuels) membres de la collection qui ne sont pas dans la liste. Une structure instantanée de la conscience est ouverte dans ce sens : si l'on se figure un certain nombre d'objets et de relations entre eux qui sont « les » objets et « les » relations de cette structure, il y aura toujours d'autres objets et relations qui sont éventuellement dans cette structure mais qui n'étaient pas pris en compte⁴³. En faisant cette liste de membres de la collection, ou en se figurant ces objets et ces relations de la structure, on *visait* à la collection ouverte, ou la structure ouverte, sans vraiment l'atteindre ; et *l'ajout* d'un membre de la collection qui était absent auparavant de la liste ne change pas le fait qu'on vise la *même* collection ouverte, de même que *l'ajout* d'un objet ou d'une relation qui ne comptait pas parmi ceux que l'on se figurait auparavant, ne change pas le fait qu'on vise toujours la même structure (ouverte) de la conscience. Si l'on relâchait l'expression en parlant de la liste des membres que l'on a déjà construite ou conçue *comme* la collection ouverte, ou des objets et des rapports figurés *comme* s'ils constituaient (la totalité de) la structure ouverte, il semblerait que l'on puisse *ajouter* aux membres d'une collection ouverte *sans* pour autant *changer* la collection, ou que l'on puisse *ajouter* des objets et des rapports à une structure ouverte *sans* pour autant *changer* la structure qu'ils constituent. À strictement parler, ceci n'est qu'une illusion produite par une imprécision dans l'expression. Néanmoins, en supposant que l'on est clair sur le sens visé de cette expression, on s'autorise à l'employer là où l'expression courte et simple est plus claire que l'expression longue et complexe, mais exacte.

Le vague de la structure instantanée de la conscience a des conséquences importantes sur les notions d'arrière-plan et de premier plan de l'état instantané de la conscience. Le caractère ouvert de la collection des objets d'une structure instantanée de la conscience à un instant⁴⁴ correspond au fait qu'il n'est pas précisément et totalement déterminé quels aspects, choses, événements, phénomènes sont *focalisés* à cet instant (§2.2.2). Or, on peut distinguer *deux* éléments de cette indétermination.

D'une part, il y a une indétermination dans les objets « nouveaux » à la structure. Par exemple, lorsque le sujet regarde sa bibliothèque, il pourrait être

⁴³Voir les remarques terminologiques de §2.1.3.

⁴⁴Le caractère vague des relations est moins intéressant ici. Il comporte un aspect en rapport avec les objets (qui ajoute des relations pour les « nouveaux objets ») et un aspect relativement indépendant des objets (un vague dans la présence même de cette relation entre les objets certainement présents).

une question indéterminée celle de savoir si tel livre au bord de son champ visuel est dans la structure instantanée de la conscience ou non, s'il y a un objet qui lui correspond. Ou, pour donner un autre exemple, il pourrait être indéterminé si, à un instant particulier de la lecture, tel personnage ou tel événement que le lecteur a rencontré ailleurs dans le livre est un objet dans la structure instantanée de la conscience. Il s'agit, pour ainsi dire, du vague « *au bord* » des objets de la structure instantanée de la conscience, et donc au « *bord* » des aspects *focalisés* dans l'état instantané de la conscience.

D'autre part, il y a une indétermination relative aux objets qui « correspondent » aux relations de la structure, c'est-à-dire une indétermination dans le statut de ces relations : telle relation est-elle explicitement remarquée (focalisée) ou non ? Pour reprendre l'exemple de la section précédente, la question de savoir si le sujet « prête l'attention » au fait que les livres sont à côté l'un de l'autre ou non peut être une question indéterminée. Dans ce genre de cas, il est question de la « ligne de démarcation » entre le focalisé et le non focalisé, entre l'objet et la (simple) relation, ligne qui est vague. Or, ce vague est apparemment différent du vague précédent. Alors que dans le cas précédent c'est la *présence* de l'élément dans l'état de la conscience qui est indéterminée ou vague, dans le cas actuel, où l'aspect est *toujours présent dans* l'état instantané de la conscience (il y a toujours la relation appropriée), c'est le *degré* auquel il est remarqué ou explicite ou *focalisé* qui est en question (le fait d'y avoir un objet qui « correspond » à la relation ou non). Il s'agit, plutôt que d'un vague « *au bord* » de l'état instantané de la conscience, d'un vague « *en profondeur* ».

Il s'ensuit que, si l'on veut parler d'un premier plan et d'un arrière-plan de la conscience, l'arrière-plan, conçu comme ce qui n'est pas au premier plan, est *structuré selon deux dimensions différentes*. Ce qui est mis au clair à l'aide d'une analogie avec le champ visuel. Il y a d'une part des aspects qui ne sont pas au premier plan parce qu'ils sont « à côté » du premier plan. Il y a d'autre part les aspects qui ne sont pas au premier plan, mais « derrière » lui. Selon l'analogie, on rapproche les premiers aspects du vague « *au bord* » de l'état instantané de la conscience, c'est-à-dire du vague relatif à la ligne entre les aspects focalisés et les aspects totalement absents de l'état instantané (entre présence comme objet et absence). À l'opposé, on rapproche les derniers aspects du vague « *en profondeur* » de l'état instantané de la conscience, du vague relatif à la ligne entre ce qui est focalisé et ce qui est non focalisé mais présent (entre présence comme objet ou présence comme relation seule). L'arrière-plan est donc constitué *aussi bien* des aspects dont l'appartenance à l'état instantané de la conscience constitue une question vague *que* des aspects qui tantôt ne sont pas focalisés tantôt ont

un statut vague par rapport à leur statut de focalisation. Il est clair, du moins, que le non focalisé, comme on l'emploie ici, ne coïncide pas avec l'arrière-plan. Ce qui explique le soin pris dans §2.2.3 pour souligner que, à la différence de certains autres conceptions, le triplet *focalisé*, *non focalisé*, et *absent* ne constitue pas une gradation simple.⁴⁵

Le caractère vague de la structure instantanée de la conscience rend le travail extrêmement complexe et difficile. Néanmoins, le vague de la structure instantanée de la conscience joue un rôle capital dans la pensée de la dynamique. Pour aborder et comprendre ce rôle, il faut prendre en compte toute la complexité du vague : ceci sera le sujet du reste de cette section. Par ailleurs, ayant précisé un certain nombre de conséquences importantes et utiles du vague (et de son rapport à la dynamique), on se permettra à plusieurs moments dans les sections et chapitres suivants de travailler en termes des objets et des relations assez précis, en supposant entendu l'élément omniprésent du vague et ses conséquences. En particulier, dans la considération des exemples particuliers, on choisira souvent une structure instantanée précise « typique », qui comporte dans un certain sens les propriétés « importantes » de la structure vague vraiment en jeu. Alors que d'autres choix pourront être possibles, les différences entre eux ne seront pas importantes pour les usages que l'on en fera.

⁴⁵ Conforme au thème de ce paragraphe, cette distinction entre la « largeur » de l'arrière-plan et sa « profondeur » est elle aussi vague. Par ailleurs, bien qu'elle puisse exercer une attraction intuitive, il faut la manier avec soin. En particulier, si, au lieu de travailler avec des structures au sens faible de §2.1.3, on impose des conditions plus exigeantes – notamment les conditions faisant référence (plus ou moins explicite) à la totalité de la structure – cette distinction risque de s'écarter. Par exemple, si l'on exige qu'il y ait une opération de complémentation sur les objets de la structure, tout ajout d'un objet correspondant à une relation, lequel est un changement selon la profondeur, exigera l'ajout de son complément, ce qui ressemble plutôt à un changement selon la largeur. Et si, de surcroît, on exige qu'il y ait la disjonction et la conjonction, tout ajout d'un objet « à côté » implique l'ajout d'un objet « en profondeur » (des conjonctions avec d'autres objets). En conséquence, cette distinction n'est pas valide sur les structures logiques, qui sont (au moins) des algèbres de Boole. Ce qui exige un soin supplémentaire dans la manipulation de leurs objets (propositions, ensembles de mondes possibles) : par exemple, toute tentative d'écarter des « mondes » ou des propositions « non intéressants » ou *hors sujet* « au bord » a des conséquences « en profondeur ». C'est d'ailleurs une des raisons pour des difficultés relativement aux phénomènes d'indépendance, de *hors sujet* et de la dynamique (§§4.2.3, 4.2.4 et 4.3.1) qui touchent aux approches des contrefactuels en termes de mondes possibles (sur le rapport entre les algèbres de Boole et les mondes possibles, voir §4.2.1).

2.3.2 La continuité de la dynamique

Où l'on introduit la continuité, c'est-à-dire la propriété de la dynamique d'un système selon laquelle les états instantanés à des moments temporels voisins sont proches ou se chevauchent. Comme telle, la notion de continuité est relative à une notion de proximité entre états instantanés, ou à une notion de chevauchement des états instantanés ; bref, elle est relative à un espace qui contient les états instantanés. La continuité pourrait donc être comprise comme une contrainte, sur la représentation de l'état instantané, qui demande qu'elle entraîne un espace relatif auquel la dynamique soit continue. Vu l'importance de la continuité pour d'autres théories ayant trait à une dynamique, on admet ici cette exigence méthodologique sur la représentation de l'état instantané de la conscience.

On a souvent remarqué que l'expérience a un certain degré de continuité. Quels que soient les virements, les surpris, les sauts dans la pensée, il reste toujours quelque pensée ou aspect gardé, même pour un moment. James affirme que « la pensée est sensiblement continue » ; Husserl, pour sa part, parle de la conscience comme un « unique « flux du vécu » », comme un « continuum rempli »⁴⁶.

Ces observations suggérant que la dynamique de la conscience est continue impliquent une contrainte sur la représentation adoptée de l'état instantané : puisque la dynamique consiste en la succession des états instantanés de la conscience, on n'a pas rendu compte de la continuité de la dynamique si la représentation des états instantanés que l'on emploie n'est pas telle que leur succession soit continue. On admet cette contrainte ici, mais non pas au nom de quelque « constat » à l'égard de la conscience. Elle est plutôt conçue comme une *contrainte méthodologique* qui s'applique *au cadre employé pour penser et représenter la conscience* : pour qu'un cadre théorique soit acceptable, il faut qu'il permette de penser la dynamique de la conscience comme continue. Cette contrainte méthodologique relève de la conjoncture de la volonté de prendre en compte la dynamique de la conscience d'une part, et par l'idée que *toute pensée vraiment féconde de la dynamique d'un système la représente comme continue* d'autre part. À l'appui de cette idée, on constate que dans la physique, sans doute la pensée la plus avancée de la dynamique, la plupart des systèmes sont représentés de telle façon que leurs dynamiques sont continues. D'ailleurs, en employant un exemple provenant de la physique, on pourra non seulement illustrer l'impor-

⁴⁶James (1981, Ch. IX) ; Husserl (1913a, p163), ou également (1913b, V^e Recherche, §1).

tance de la continuité, mais également approfondir son sens.

Il s'agit d'un navire qui se déplace sur la mer. On pourrait se figurer les états instantanés du navire comme des « régions » qu'il occupe dans la mer ; ainsi, le mouvement du navire correspond à une succession de telles « régions ». Le mouvement du navire est *continu* dans la mesure où la région occupée à un moment se chevauche avec celle occupée un second plus tard. Ce chevauchement constitue un rapport entre l'état instantané du navire et sa dynamique, rapport que l'on identifie comme indice d'une représentation féconde du navire et de son mouvement, rapport que l'on nomme *continuité*.

Pour approfondir un peu la question, on constate que cette notion de continuité est relative à un espace, à savoir l'espace constitué par la mer. À strictement parler, on peut traiter des états instantanés du navire comme des entités abstraites sans rapport entre elles. Il est nécessaire de les ramener à un *espace* commun pour qu'ils puissent être considérés comme des *régions* et pour qu'il fasse sens de parler de *chevauchement* entre eux. Ils sont des régions seulement relatives à cet espace ; et leur dynamique est continue seulement relative à cet espace. Pourtant, il ne faut pas entendre là quelconque primauté du terme espace, car ces énoncés peuvent être également lus en sens inverse. Un *espace* livre une notion de *région* et de *chevauchement* : à l'inverse, une notion de *région* munie d'une notion de *chevauchement* des régions implique un *espace*⁴⁷. Bien entendu, pour des notions inouïes de région et de chevauchement, cet espace serait abstrait, mais néanmoins un espace.

Le point capital est le suivant : le fait de bien représenter la dynamique du navire consiste à avoir trouvé une notion d'espace, de région et de chevauchement qui « accueillent » les états instantanés et représentent leur dynamique comme continue. Puisque dans ce cas physique le bon espace est immédiat, on a tendance à oublier son importance ; néanmoins, il s'avère être le pierre angulaire de l'analyse de la déplacement du navire. Toutes les théories physiques qui portent sur sa dynamique se formulent non seulement en termes de cet espace, mais de plus s'appuient sur des notions, telles vitesse, moment, force, qui se définissent à partir de ou en se rapportant à cet espace. Pour ainsi dire, un espace relatif auquel la dynamique du système est continu est une *condition*

⁴⁷Étant donné, bien entendu, quelques contraintes sur la région et le chevauchement. On entend l'espace ici dans le sens de la topologie, discipline mathématique établie par Hausdorff (1914) et Kuratowski (1966) visant à étudier les propriétés de l'espace au sens le plus abstrait. Selon la topologie, un espace (topologique) est un ensemble de « ensembles ouverts » muni des opérateurs d'intersection et d'union. Dans le cas discuté, la région est l'ensemble ouvert, le chevauchement est l'intersection ; corrélativement, les contraintes dont on parle sont les axiomes de la topologie.

préalable au développement d'une théorie de la dynamique du système.

Par conséquent, on admet le principe méthodologique selon laquelle on juge *toute proposition de représentation d'état instantané et toute proposition d'espace, de région ou de chevauchement* qui l'accompagne selon la possibilité qu'elles offrent de *penser la dynamique comme continue*.

Il reste néanmoins que, en raison des liens logiques entre les notions d'espace, de région et de chevauchement, il y a deux manières à procéder pour aborder le problème de la dynamique. D'une part, on peut chercher *directement l'espace approprié*, lequel entraîne les notions de région et de chevauchement. D'autre part, on peut chercher *d'abord les bonnes notions de région et de chevauchement*, lesquels impliquent un espace.

Dans le cas de la physique, où le bon espace est donné immédiatement, la première approche est la plus naturelle, et d'ailleurs celle qui est prise. En revanche, dans le cas de la conscience, il n'y a pas à première vue d'espace immédiat et naturel pour étayer une bonne représentation de la dynamique. Il faut par conséquent approcher du problème de la dynamique de l'autre bout, en cherchant une notion de région et une notion de chevauchement appropriées à la dynamique des états instantanés de la conscience. De fait, on s'efforcera de penser les états instantanés de la conscience comme des *régions* locales, dont ceux qui sont adjacentes dans la dynamique *se chevauchent*. Si c'est possible, on aura pensé la dynamique des états instantanés de la conscience comme continue, et ainsi fait un grand pas vers la prise en charge de la dynamique et de son rapport étroit aux états instantanés.

Dans le paragraphe suivant, on pointe les notions de région et de chevauchement appropriées à penser la dynamique de la conscience, en les esquissant à grands traits sans s'attarder sur les détails complexes. L'apport clef à la conception de telles notions de région et de chevauchement est le vague de l'état instantané de la conscience.

2.3.3 La continuité et le vague

Où l'on conçoit une notion de continuité sur les successions des structures instantanées de la conscience, telle que la dynamique de la conscience est continue en ce sens. Selon une définition abstraite de la continuité, la dynamique est continue si les aires vagues appropriées à des moments temporels proches se chevauchent. Dans la conception proposée de la dynamique de la conscience, les aires vagues sont les structures instantanées de la conscience, et la notion

de chevauchement est de l'ordre d'une correspondance entre certains aspects de différentes structures : la dynamique est continue dans la mesure où les structures instantanées à des moments prochains ont des aspects qui sont en correspondance les uns avec les autres. L'approche repose donc sur un lien entre le vague et la continuité : le vague de la structure instantanée lui permet les chevauchements avec d'autres structures instantanées qui constituent la continuité de la dynamique.

Une condition préalable sur toute pensée féconde de la dynamique, et donc de la dynamique de la conscience, est de la concevoir comme continue. On arrive à proposer une conception de la dynamique comme continue, laquelle s'appuie sur des structures instantanées de la conscience développées ci-dessus, par l'intermédiaire d'une considération du rapport entre le vague et la dynamique. Une analogie, toujours avec la physique, permet de bien saisir ce rapport et de bien apprécier l'approche de la dynamique de la conscience qui sera présentée ci-dessus.

On s'imagine regarder un objet physique en mouvement (un oiseau en vol, par exemple), perçu comme un *point* en mouvement (il s'agit dans cet exemple non pas des régions, comme auparavant, mais des points). Ce mouvement est, du moins intuitivement, continu. Or, si l'on essaie de saisir la position de cet objet (de ce point) dans la perception à un moment particulier, on trouve que c'est pratiquement impossible : le mieux que l'on peut faire est de repérer une aire approximative (ou une partie approximative de sa trajectoire). Autrement dit, une aire vague. Comme si, pour parler de manière relâchée, le point était vague, et, de plus, comme si son vague « répondait » à sa dynamique. Muni de cette aire vague, il y a deux démarches possibles pour en faire une théorie précise.

D'une part, on peut *postuler* un espace de positions précises, dont l'objet occupe une et une seule à chaque moment (précis), mais que l'on ne pourrait pas repérer le point avec cette précision, et donc on « confond » plusieurs positions successives pour en faire l'aire vague. Selon ce schéma, la *continuité* du trajectoire de l'objet est conséquence de la *structure* de l'espace des positions ; notamment, cette structure comporte des « mesures » de proximité selon lesquelles les positions sont plus ou moins proches (ce qui impliquent une topologie au sens abstrait). La trajectoire est continue puisque, pour des moments temporels de plus en plus proches, les positions de l'objet à ces moments sont de plus en plus proches. En l'occurrence, l'espace de ces positions est l'espace

physique (cartésien)⁴⁸.

D'autre part, on pourrait s'efforcer de comprendre les propriétés des « aires vagues » et les rapports entre elles *en tant que telles* ; c'est-à-dire la structure *des* « aires vagues ». Selon cette approche plus « innocente », la *continuité* correspond à une *structure* des aires vagues impliquées dans le mouvement de l'objet, en l'occurrence, au *chevauchement* des aires vagues relatives à des moments proches. Ici, la *topologie* ne dépend pas d'un espace de points, mais consiste directement en *une structure sur les aires vagues*. Normalement on dit que ces aires vagues sont *ouvertes*⁴⁹.

Dans le cas de la physique – la trajectoire des objets par exemple – ces deux démarches sont, moyennant certaines hypothèses sur l'existence des aires vagues de plus et plus petites (et donc précises)⁵⁰, effectivement équivalentes. Toutefois, la différence d'approche est significative : l'une présuppose ou pose l'espace d'abord ; l'autre trouve l'espace après. Notamment, la première approche implique une double thèse. Premièrement, elle suppose qu'*il y a* un ensemble de positions précises qui constituent l'espace, alors qu'il pourrait ne pas être toujours évident que la structure des aires ouvertes (ou vagues) en permet un⁵¹. Elle soutient deuxièmement que l'ensemble de positions posé *est bien* l'espace des aires vagues, ce qui, étant donné la difficulté de ramener les aires vagues aux collections de positions précises, n'est pas nécessairement évident. En revanche, la dernière approche est plus sûre, même si peut-être moins intuitive.

Il est temps de revenir à la dynamique des états instantanés de la conscience. On a remarqué la difficulté de trouver ou de penser un espace qui permettra de penser la dynamique, et notamment une dynamique continue. Mieux vaut être prudent et, en suivant l'analogie physique, s'efforcer de comprendre la structure *des* aires ouvertes : si l'on peut dire, la structure *sur* des aires ouvertes. Il est désormais clair que les analogues des aires vagues sont les structures instantanées

⁴⁸La notion de continuité comme proximité agrandissant des points correspondant à des instants de plus en plus proches dans un espace donné est la notion reconnue dans l'analyse (mathématique) ; pourtant, elle peut être ramené à la notion de continuité discutée ci-dessus (en termes des intersections des régions), alors que l'inverse n'est pas nécessairement vrai.

⁴⁹Alors que classiquement (Kuratowski, 1966) un espace topologique est défini comme une structure sur un ensemble de points, il y a une définition équivalente en termes seulement des ensembles ouverts (ou aires vagues) qui ne présupposent pas des points, ce qui inspire la dite « topologie formelle » ; voir par exemple Sambin (2003). Voir la note 47 sur la topologie.

⁵⁰Précisément : des chaînes infinies des ensembles ouverts continus l'un dans le précédent. Voir par exemple Whitehead (1920, Ch IV, surtout p79 *sq.*), qui aborde la question à partir des « événements » au lieu des « aires ouvertes ».

⁵¹Par exemple, dans les cas où l'hypothèse des aires ouvertes de plus en plus petites n'est pas justifiée.

de la conscience, lesquelles sont *vagues*. D'où deux conclusions de l'analogie.

En premier lieu, de même qu'il y a un rapport entre le vague des « aires » et la dynamique du mobile, de même il y a un rapport entre le caractère ouvert des structures instantanées de la conscience et la dynamique de ces structures, externe à chacune d'elle. Ce caractère ouvert donne un sens à la continuité de cette dynamique. Elle est continue dans la mesure où les structures instantanées ouvertes à des moments proches *se chevauchent*. L'explicitation du sens précis de ce chevauchement entre structures instantanées ouvertes appartient à une étude détaillée de la dynamique, laquelle déborde les limites de cette thèse⁵². Il ne s'agit ici que d'exposer l'*idée* et de souligner le rapport entre la continuité et le vague. Il faut tenir compte du vague des structures instantanées pour pouvoir les penser comme des régions qui peuvent se chevaucher avec d'autres régions (structures instantanées ouvertes), et pour ainsi constituer une pensée de la dynamique.

Dans la suite, on se contente avec une notion intuitive de chevauchement entre structures instantanées de la conscience. À cet égard, l'intuition la plus naturelle consiste à considérer le chevauchement comme une forme de communauté d'objets et de relations entre les structures (une « intersection », si l'on veut), c'est-à-dire comme se rapportant à une « correspondance » ou une « identification » des aspects des différentes structures entre eux. En conséquence, la « correspondance » entre différentes structures entretient un rapport étroit avec la question de la dynamique⁵³. Pour le propos de ce travail, dont le but principal est d'aborder la question de la fiction, on se permet d'éviter les complexités de la « correspondance », en parlant comme si l'on disposait d'une telle « correspondance » (§2.2.1). Avec cette notion intuitive de chevauchement comme « correspondance » entre des aspects, une succession des structures instantanées de la conscience est continue si des structures instantanées à des instants voisins ont des aspects qui « correspondent » les uns aux autres. Or, cela semble être bien le cas, surtout étant donné le *vague des structures instantanées*, et donc de *l'appartenance des aspects aux structures*. Le vague des structures instantanées non seulement permet de les penser comme des régions, mais comme des régions qui se chevauchent ; autrement dit, il permet de penser leur dynamique comme continue.

⁵²Quelques pistes techniques sont discutées, dans le cas particulier des algèbres de Boole, dans la seconde partie de la thèse (notamment §§5.1.5 et 5.2.3) et dans l'Annexe (notamment §A.3).

⁵³Comme on verra dans §2.4, ce rapport étroit est important pour établir la « correspondance » entre structures instantanées, dans la mesure où cette « correspondance » doit respecter la contrainte imposée par la continuité.

La deuxième conclusion qui s'impose à la suite de l'analogie avec la physique concerne la décision, prise dans §2.3.1, de prendre en charge le vague. En mettant en évidence le rapport entre vague et dynamique, l'analogie montre que la prise en charge du vague implique *premièrement* une prise en charge de la question de la dynamique des états instantanés de la conscience et *deuxièmement* l'adoption d'une certaine stratégie pour l'aborder, à savoir celle, plus prudente, qui étudie les structures *sur* les états instantanés vagues, sans poser un espace « d'états instantanés *précis* » ni se soucier de la question de « quel état instantané précis » est « occupé » par la conscience à un instant donné. Si, par opposition, on décide de rabattre la structure ouverte de la conscience à un instant sur quelconque structure précise (voir la discussion de Fine et de Williamson dans §2.3.1), de deux choses l'une. Tantôt on s'occupe du problème de se débarrasser du vague, sans se soucier de la question de la dynamique, de telle sorte que, quant à cette dernière question, la structure précise proposée n'est même pas soutenue être apte ou utile. Tantôt on entend dans la position d'une structure précise la *préconisation* d'un espace comme approprié à la pensée de la dynamique ; or, étant donné le manque absolu de soutien pour une telle suggestion et de discussion d'elle, et étant donné l'apparence maladroite de telles structures précises, laquelle aurait sans doute des conséquences nuisibles sur leurs emplois dans les considérations des questions dynamiques, cet espace préconisé ne semble pas convenir. En résumé, tantôt on parle d'une question différente et *hors sujet* ici, tantôt on soutient une thèse relative à la question qui importe ici, mais la thèse est carrément sans justification (et probablement fausse). En tout cas, comme l'a été anticipé dans §2.3.1, l'approche qui ramène le vague au précis n'est pas féconde par rapport à la question de la dynamique posée ici.

On a mis l'accent auparavant sur l'importance d'une dynamique pensée proprement, c'est-à-dire de manière continue, pour mettre en rapport les états instantanés et leur succession dynamique. Dans le cas présent, les états instantanés sont les structures instantanées *ouvertes* de la conscience, le rapport continu entre elles est le *chevauchement* des structures ouvertes successives. Ce rapport étroit entre la structure instantanée ouverte individuelle et les autres structures qui le succèdent ou qui le précèdent a quelques conséquences significatives pour la suite.

2.3.4 Rapport entre la structure instantanée de la conscience et sa dynamique

Où l'on explicite deux conséquences d'avoir représenté la dynamique de la conscience comme continue, c'est-à-dire, de l'avoir représentée de telle façon qu'il y a un rapport étroit entre la structure instantanée de la conscience à un instant et les structures à des instants précédents et suivants. D'un côté, certains aspects de la dynamique de la conscience peuvent être compris aussi bien comme synchroniques, ayant trait à une structure instantanée à un instant seul, que comme diachroniques, référant à une succession de structures instantanées. De l'autre côté, certaines propriétés d'une structure instantanée peuvent entraîner des effets sur les propriétés des structures instantanées à des instants voisins.

Dans un système dynamique pensé de manière continue, l'état instantané importe pour la dynamique, dans la mesure où elle pose une contrainte sur les états instantanés à des instants immédiatement ultérieurs (et antérieurs). Dans le cas de la conscience, c'est-à-dire que la structure instantanée de la conscience importe relative à la dynamique des structures instantanées. Dans la mesure où l'on conçoit la structure instantanée ouverte comme *région*, et le rapport avec d'autres structures antérieures ou ultérieures comme *chevauchement*, le rapport étroit entre l'état instantané (la structure instantanée de la conscience) et sa dynamique consiste effectivement en un rapport entre un aspect de *l'intérieur* de la structure instantanée de la conscience et des questions qui ne se posent qu'à *l'extérieur* de la structure instantanée de la conscience – quelles structures suivent ou précèdent celle-ci (ou quelles propriétés ont-elles) ?

Ce rapport est d'une double importance ici. Premièrement, il permet d'alléger l'expression et de focaliser l'enquête, car il permet parfois d'aborder des aspects de la dynamique d'une succession de structures instantanées de la conscience – des questions *externes* – en considérant des aspects de certaines structures particulières de cette succession – des questions *internes*. Il ne s'agit nullement d'un artifice qui s'appuie sur la continuité, mais une propriété d'importance cruciale que l'on a tendance à sous-estimer : puisque l'état instantané (représenté correctement) reflète des aspects dynamiques et que la dynamique est conséquence des états instantanés, il ne fait pas toujours du sens d'insister pour qu'un aspect soit diachronique (relatif à la succession) plutôt que synchronique (relatif à un état instantané particulier). Malgré les multiples débats et discussions, il n'y a parfois aucun fait (« *fact of the matter* »). D'où l'on peut aborder

l'aspect *tantôt* dans son versant diachronique (comme aspect de la succession) *tantôt* dans son versant synchronique (comme aspect d'un ou des état(s) instantané(s)) sans prendre partie dans ces débats vides. Ceci vaut particulièrement pour des aspects dynamiques de courte durée (relatifs aux successions courtes), car la continuité implique que la mise en rapport des structures instantanées immédiatement voisines dans le temps.

On tire profit de cette possibilité dans les discussions ultérieures, en abordant parfois des transitions particulières entre structures instantanées de la conscience à des instants successifs en repérant certaines structures instantanées qui comportent (ou, dans la mesure où elles sont vagues, *peuvent* comporter⁵⁴) les deux termes impliqués dans la transition. C'est-à-dire, on aborde la transition en considérant la (les) structure(s) qui est (sont), pour ainsi dire, au « *pivot* » de la transition. Dans la mesure où les structures antérieures et ultérieures sont continues avec celle-ci, c'est-à-dire lui chevauche, des propriétés de la transition (dynamique) des premières aux dernières « se reflètent » comme aspects (instantanés) de cette structure instantanée. On pourrait aborder ainsi des aspects de la dynamique dans le cadre plus maniable d'une seule structure instantanée (à un instant).

Deuxièmement, ce rapport étroit entre aspects *internes* et aspects *externes* à la structure instantanée de la conscience entraîne la question des *effets* ou des *influences* ou du moins *de la signification* des objets, relations ou aspects d'une structure instantanée sur ou pour la dynamique. Si la continuité de la dynamique implique un rapport entre les aspects de la structure instantanée de la conscience à un instant et ceux des structures instantanées suivantes (et précédentes), le fait pour la première structure de comporter certains objets, relations ou aspects peut exercer des influences voire des contraintes sur les propriétés des dernières structures, et on pourrait vouloir caractériser les différents genres d'effets ou de contraintes impliqués par les différents aspects. Par exemple, pour un aspect au premier plan de l'état instantané de la conscience, c'est-à-dire un objet (appartenant pleinement à la structure instantanée de la conscience), il faut que les structures suivantes comportent cet aspect, qu'il soit au premier plan, ou qu'il s'évanouisse dans l'arrière-plan, tantôt comme aspect focalisé mais « au bord » de la conscience, tantôt comme aspect non focalisé (« en profondeur »

⁵⁴Le vague ou le caractère ouvert de la structure est important ici : il permet de travailler dans l'hypothèse que les deux termes sont dans la même structure. Pour ceux qui connaissent un peu la topologie, plus « l'instant temporel » est grossier (comme « intervalle » sur le continu du temps), plus la région ouverte de la structure instantanée à cet « instant » le sera ; d'où la possibilité de considérer dans une même structure ouverte deux éléments qui pourraient éventuellement être distingués avec une échelle temporelle plus précis.

de la conscience ; voir §2.3.1). Que l'on considère, en guise d'illustration de ces possibilités, un instant où le sujet lit « Holmes ouvrit la porte » – les deux objets *Holmes* et *porte* sont au premier plan dans la structure instantanée de la conscience à cet instant – suivi par un instant où il lit « Il était en face d'un escalier, qu'il commença à monter ». L'objet *Holmes* est semble-t-il retenu au premier plan dans les structures instantanées de la conscience aux instants suivants. L'objet *porte* s'en éloigne de plus en plus ; pourtant, il ne disparaît pas de la structure instantanée de la conscience au prochain instant, en y demeurant toujours objet, mais, dans la collection ouverte des objets, son appartenance devient de plus en plus douteuse⁵⁵. Or, alors que l'objet *porte* s'évanouit comme objet, elle pourrait demeurer sous forme de relation : par exemple, elle peut jouer un rôle dans le fait de « connaître » la position de Holmes (sur l'escalier, à l'intérieur d'un bâtiment), connaissance qui est probablement non focalisée, en se manifestant à travers les propriétés des relations (de position entre le Holmes avançant et la sécurité du dehors)⁵⁶. Le premier cas est un exemple d'un aspect au premier plan qui demeure au premier plan, les deuxième cas est un exemple d'un aspect au premier plan qui tombe dans l'arrière-plan tout en restant focalisé, et le dernier cas est un exemple d'un aspect au premier plan qui tombe dans l'arrière-plan en devenant non focalisé.

Plus intéressante que les effets dynamiques des objets (ou des relations) particuliers, est l'importance pour la dynamique des aspects « globaux » de la structure instantanée. C'est-à-dire des aspects de la structure instantanée en tant que *totalité*, la « forme » de la structure, si l'on peut dire. Car, s'il y a chevauchement entre une structure instantanée particulière et les structures instantanées ultérieures, alors la forme de la structure, ou ses aspects d'ensemble, a un effet sur les objets et les relations qui figurent dans les structures ultérieures. Par exemple, si la structure instantanée de la conscience à un moment ne comporte peu d'autre qu'un objet particulier, si, pour ainsi dire, la structure est *concentrée* sur cet objet, cela impose une forte contrainte sur les structures instantanées qui peuvent suivre. Il n'y a pas, hors cet objet, beaucoup d'autres aspects de cette structure qui peuvent étayer un chevauchement, en telle sorte que le sujet continuera (tout probablement) de s'en soucier. Cette importance de la *totalité* de la structure instantanée pour la dynamique a plu-

⁵⁵ Cf. la rétention de Husserl (1928, §11 *sq.*). Or, contrairement à l'analyse husserlienne, on n'admet pas ici de différence de genre entre l'objet présent (la « conscience impressionnelle ») et l'objet passé (la « conscience rétentionnelle »). Voir aussi la note 73 de §2.5.1.

⁵⁶ *En gros*, la situation est pareille pour une relation d'une structure instantanée : elle pourrait rester non focalisée, s'évanouir des structures instantanées au fur et à mesure que le temps passe, ou devenir focalisée (par exemple, lorsqu'elle est « remarquée » par le sujet).

sieurs conséquences intéressantes. D'une part, elle implique une importance de l'arrière-plan de la structure, lequel, en raison de son caractère fugace (dès qu'on souligne un aspect de l'arrière-plan, il n'est plus à l'arrière-plan) ne pourrait être saisi ou représenté de manière féconde que dans la totalité de la structure. Par exemple, la *concentration* de la structure sur l'objet semble correspondre à ce qu'on qualifie d'une *étroitesse* de l'arrière-plan de cette structure. D'autre part, puisque, comme on le verra dans §2.5.2, les aspects globaux de la structure instantanée de la conscience se rapportent aux sentiments du sujet, les effets dynamiques de tels aspects jouent un rôle important dans la compréhension de ces sentiments.

2.3.5 Résumé : dynamique, continuité et vague

La question difficile et épineuse de la fiction, et de son rapport à la réalité, a été reconnue dans Chapitre 1 comme celle de sa dynamique (pour autant qu'elle se pose relative à la psychologie du lecteur). Le recours à la conscience, argué dans §§1.4 et 1.5 n'est utile qu'à condition qu'il soit possible de penser la dynamique en termes d'elle. Il faut donc prendre en charge la question de la dynamique de la conscience.

Une condition préalable sur une approche féconde de la conscience, capable d'étayer une pensée de la dynamique est la *continuité* de la dynamique dans cette approche. Car la continuité exprime le rapport entre l'état instantané et sa dynamique, ou, plus précisément, les rapports entre les états instantanés successifs dans la dynamique. La continuité n'est donc pas conçue comme une propriété repérée dans la succession d'états instantanés de la conscience, elle est au contraire une exigence méthodologique sur la représentation de cette succession et de ces états. Si une représentation des états instantanés d'un système (la conscience par exemple) n'admet pas une notion d'espace⁵⁷, ou si elle implique une notion d'espace telle que la succession des états instantanés dans la dynamique est discontinue relative à cet espace, alors elle est insuffisante à penser la dynamique.

Dans le cas de la conscience, l'apport clef pour permettre la pensée de la dynamique est le caractère *ouvert* ou *vague* de l'état instantané de la conscience. Dans la représentation de la conscience proposée ici, la structure de l'état instantané de la conscience est *ouverte*. Dès lors, les structures instantanées de la conscience sont considérées comme des régions vagues, et la continuité est conçue en termes du chevauchement entre les structures instantanées à des ins-

⁵⁷Ceci n'arrive jamais en pratique : normalement il y a un espace, mais un espace d'une intuition faible.

tants successifs. Il ne s'agit nullement de « préciser » ces structures, ni de poser d'un espace, mais d'aborder la continuité en termes des rapports appropriés entre les structures instantanées ouvertes. Bien qu'on n'aborde pas cette question dans ces détails techniques et complexes dans ce travail, il devrait être clair que l'on a pris conscience des complexités de la question de la dynamique, et que le cadre proposé comporte la possibilité, voire la promesse, de l'aborder au fond.

Enfin, la continuité de la dynamique répond à une exigence de mettre en rapport l'état instantané et sa dynamique ; en l'occurrence, il y a un rapport important entre l'*intérieur* de la structure instantanée – ses objets, ses relations, ses aspects globaux – et son *extérieur* – les structures qui le suivent ou le précèdent dans la dynamique. À part quelques conséquences pratiques, dont une compréhension de l'importance de l'arrière-plan de la structure instantanée de la conscience, ce point était une méthode qui sera employée à plusieurs reprises dans la suite. À savoir, il permet dans plusieurs cas d'aborder des aspects dynamiques ou des transitions dynamiques – c'est-à-dire des caractéristiques *des successions* de structures instantanées de la conscience – *dans* une des structures instantanées.

2.4 Se servir des structures instantanées de la conscience : le rôle de l'interprétation

Il convient à présent de faire le point de la discussion. Les données dans le cadre « internaliste » adopté ici consistent en une succession d'états instantanés de la conscience. Elles peuvent être représentées d'une manière rigoureuse par des structures, munies des objets et des relations, au sens de §2.1.3. On a vu que ces structures instantanées de la conscience sont locales et vagues, et on a identifié la distinction entre objet et relation à celle entre les aspects focalisés et non focalisés dans l'état de la conscience.

À partir de cet ensemble de données, il y a deux types d'opérations que l'on pourrait vouloir effectuer. D'une part, on pourrait s'efforcer de « comprendre » cette succession comme telle, en repérant *directement* des « régularités ». Ainsi, on pourrait construire *une théorie de la dynamique des états instantanés de la conscience*⁵⁸. Dans §2.3, on a identifié comme une condition préalable à

⁵⁸Bien entendu, il y a une limite des régularités qui peuvent être trouvées, laquelle est conséquence du fait que, dans cette optique, la distinction entre les aspects « venant du monde » et les aspects « propres à la psychologie du sujet » n'est pas donnée immédiatement (si l'on s'y intéresse, elle est à trouver dans les structures instantanées et les rapports entre

une théorie de la dynamique le fait de la représenter comme continue, et on a jeté les bases d'une telle approche « directe », en concevant les états instantanés de la conscience de telle manière que leur dynamique soit effectivement continue. Il ne s'agit pas, dans ce travail, d'aller plus loin dans le développement d'une théorie de la dynamique. D'ailleurs, comme on le verra dans §§3.1 et 3.2, les outils et les concepts proposés dans les trois sections précédentes (§§2.1–2.3) suffisent déjà pour permettre une première approche des questions complexes relatives à la fiction.

D'autre part, il est un autre travail, requérant d'autres opérations, de faire le rapport entre la succession d'états instantanés de la conscience du sujet et les États Mentaux que l'on lui attribue. Ce travail sera entamé, pour ce qui concerne les États Mentaux en général, dans la section suivante, la dernière de ce chapitre, où l'on montre comment attribuer des États Mentaux à un sujet sur la base de la succession d'états instantanés de la conscience. Comme on l'a déjà remarqué, ce travail n'est pas nécessaire pour comprendre les complexités de la fiction, dans la mesure où elles seront abordées au niveau de la conscience, en s'appuyant sur les outils développés dans §§2.1–2.3, sans faire entrer les États Mentaux. Néanmoins, on consacrera une partie de l'analyse de la fiction aux États Mentaux proposés comme spécifiques à la fiction (§3.3), pour montrer que l'on peut les retrouver dans le cadre psychologique développé, et d'ailleurs que le fait de les considérer dans ce cadre met au clair les raisons pour leurs faiblesses ; ce travail se servira des analyses des États Mentaux en général effectuées dans §2.5.

Dans chacune de ces deux opérations, pour chacune de ces enquêtes partant de la succession des états instantanés de la conscience, l'*interprétation* de ces structures instantanées joue un rôle crucial. En deux sens : premièrement, l'interprétation des structures instantanées implique une « identification » ou du moins une « correspondance » entre les aspects des structures différentes, laquelle est une condition préalable et un aspect fondamental de toute considération des rapports, dont les rapports dynamiques, entre eux. Deuxièmement, toute attribution d'un État Mental au sujet sur la base de ses structures instantanées présuppose l'identification des aspects de ces structures avec des concepts de celui qui attribue, dans la mesure où l'État Mental attribué est formulé à son langage à lui ; autrement dit, il faut une interprétation des structures instantanées dans le langage de « l'attributeur ». Il s'agit, dans cette courte section, de s'attarder sur la question de l'interprétation, pour clarifier son rôle et son statut dans cette thèse.

elles).

Il va de soi que toute analyse portant sur ou référant à la succession des structures instantanées de la conscience a besoin des rapports entre les structures différentes, et de plus aux concepts du théoricien. À la base, ce rapport consiste en la « correspondance », ou si l'on préfère « l'identification » des objets, relations ou aspects entre des structures différentes. Cette « correspondance » s'est avérée cruciale pour la question de la dynamique, dans la mesure où la continuité de la dynamique repose sur le chevauchement entre structures instantanées de la conscience, conçu naturellement comme une certaine communauté d'objets, de relations ou d'aspects entre ces structures (§2.3.3). D'un autre côté, il s'avérera crucial pour la question des attributions des États Mentaux au sujet, dans la mesure où, pour attribuer la pensée que X ou la croyance que X (par exemple) à un sujet sur la base des structures instantanées de la conscience, il faut « identifier » quelques aspects de ces structures comme ayant affaire à X ; en d'autres termes, il faut les « mettre en correspondance » avec le terme X du langage de celui qui attribue⁵⁹.

Or, comme l'a été souligné dans §2.2.1, la question de la « correspondance » entre des aspects d'une structure instantanée et des aspects qui lui sont externes est rendue complexe par le fait qu'elle est une *totalité par rapport à elle-même* : il ne fait strictement pas de sens, dans cette structure, de parler d'un aspect qui n'y figure pas. La « correspondance » avec des aspects externes, et donc avec des autres structures instantanées doit donc être établies de l'extérieur. C'est une des tâches remplies par l'*interprétation*, dans la mesure où elle interprète des aspects (objets, relations) d'une structure comme étant la « *même* » que des aspects d'autres structures, ou des concepts de l'attributeur.

L'importance de l'interprétation des données est d'ailleurs bien connue dans la philosophie des sciences. En effet, non seulement une interprétation des données est *nécessaire* pour une théorie scientifique, dans la mesure où cette théorie ne peut s'appliquer qu'à des éléments mis dans une forme convenable⁶⁰, mais elle est également *supposée* par la théorie, dans la mesure où il n'y a pas de collection

⁵⁹L'interprétation des aspects des structures instantanées qui se sert d'un rapport entre le sujet et le monde pourraient enfreindre les contraintes « internaliste » adoptées dans ce travail (§1.1.2). Or, puisqu'il s'agit de les interpréter pour attribuer des États Mentaux, et puisque, selon certains philosophes au moins, les États Mentaux réfèrent par définition aux aspects mondains, cette transgression doit être admise comme conséquence de la question même de l'attribution des États Mentaux sur la base de la conscience. On remarque que la question de savoir dans quelle mesure l'État Mental attribué est « externaliste » ou « internaliste », au sens du débat sur le contenu des États Mentaux dans la philosophie de l'esprit (voir la note 132 de §1.5.4), renvoie aux modalités de l'interprétation des structures instantanées.

⁶⁰La nécessité d'une interprétation des données, thème kantien (1781), est fameusement soulignée par Kuhn (1970, Ch. IX).

de données « nues », sans interprétation aucune, qui est d'abord ramassée pour ensuite être transformée dans une forme convenable à la théorie⁶¹. De même que dans les sciences, non seulement l'interprétation des objets, relations et aspects des structures instantanées est nécessaire pour parler d'eux, mais on se permet de *supposer* de telles interprétations, au fur et à mesure qu'elles sont requises⁶².

Deux remarques s'imposent à l'égard des interprétations supposées dans ce travail. Premièrement, il n'y a normalement pas une unique interprétation qui pourrait être légitimement supposée; au contraire, il y a souvent plusieurs interprétations crédibles. Pour reprendre l'exemple de §2.2.2, la structure instantanée où le fait que les deux livres sont adjacents est *focalisé*, que l'on a interprété comme comportant deux objets *livres* et un objet *adjacence* avec une relation « ... et ... entrent dans la relation de ... », pourrait également être interprétée comme comportant les *livres*, un objet *les deux livres sont adjacents*⁶³ et une relation « ... réfère à ... et ... ». De plus, les manières de concevoir la structure instantanée à l'œuvre à un moment et dans une situation particulière se multiplient si l'on ajoute le caractère vague de la structure (§2.3.1). Pour les exemples considérés dans ce travail, les différences entre les diverses possibilités de structure et d'interprétation qui permettent une analyse de l'exemple ne sont pas importantes, de telle sorte que la discussion que l'on fait en s'appuyant sur une structure et une interprétation se répéterait si l'on prenait une autre structure et une autre interprétation.

Deuxièmement, dans tous les exemples considérés ici il ne s'agit que des structures instantanées de la conscience à des instants appartenant à un *intervalle limité de temps*, de sorte que seulement l'interprétation d'un nombre restreint de structures instantanées n'est requise. De fait, *toutes les interprétations* employées et supposées ici sont *locales*, en ce sens qu'elles s'appliquent à certaines structures instantanées de la conscience à certains instants seulement, et n'affirment rien sur les structures instantanées de la conscience à d'autres instants⁶⁴. En se servant des interprétations locales, et surtout des interprétations

⁶¹ « Les paradigmes déterminent en même temps de grands domaines de l'expérience », Kuhn, *op cit.*, p180.

⁶²L'analogie avec la science qui soutient la supposition des interprétation repose sur le fait de pouvoir considérer les différentes enquêtes reposant sur les structures instantanées de la conscience (analyse de la dynamique, attribution des États Mentaux) comme des projets de constitution de quelque chose de l'ordre d'une théorie scientifique (pour le cas des États Mentaux, voir §1.3.2).

⁶³Rappeler les réserves terminologiques à l'égard du terme objet dans §2.1.3.

⁶⁴Dans la mesure où les structures instantanées dans la base d'une attribution des États Mentaux doivent être interprétées dans cette attribution, le sens de la *localité* de l'interprétation employé ici se rapproche du sens de *localité* de la base de l'attribution des États

qui ont trait à des instants successifs, la supposition d'une interprétation donnée devient moins audacieuse, et cela pour deux raisons. D'une part, comme on l'a remarqué dans §2.2.1, la difficulté dans la « correspondance » des aspects (objets, relations) entre structures instantanées se tempère au fur et à mesure que les structures en jeu s'approchent dans le temps. Alors que cette difficulté se pose de manière la plus grave relativement aux « correspondances » entre structures instantanées éloignées dans le temps (*cf.* l'exemple du *Notre Dame* vu par le même individu en enfant et en adulte, §2.2.1), il semble y avoir une « identification naturelle » entre les structures à des instants voisins (deux structures instantanées de la conscience à des instants voisins ont « intuitivement » beaucoup d'objets et d'aspects en commun). Cette identification permet d'étayer l'interprétation de ces structures, dans la mesure où celle-ci respecte celle-là. D'ailleurs, et d'autre part, toute interprétation *doit* rendre compte de ce genre d'identification entre structures à des instants voisins, dans la mesure où elle est soumise à l'exigence de *continuité*. Cette exigence impose une contrainte sur l'ensemble de notions employées (structure instantanée, chevauchement, et ainsi de suite), selon laquelle des structures instantanées de la conscience à des instants successifs se chevauchent (§2.3, notamment §§2.3.2 et 2.3.3). Or, dans la mesure où ce chevauchement est conçu comme une « correspondance » ou une « identification » entre des aspects des structures instantanées, et donc est renvoyé finalement à l'interprétation des structures, la continuité requiert, *de la part de l'interprétation supposée*, qu'elle identifie des aspects des structures instantanées qui se côtoient dans le temps.

Ces réserves et complexités étant notées, on peut aborder la question de l'attribution des États Mentaux sur la base de la succession des structures instantanées de la conscience. C'est le sujet de la section suivante.

2.5 Attribuer à partir des états instantanés de la conscience

Dans le Chapitre 1, on a soutenu que les États Mentaux quotidiens sont *nettement insuffisantes* pour aborder les questions épineuses de la fiction (en tant qu'elles se posent relativement à la psychologie du lecteur). Ces considérations ont donné lieu de penser que ces questions peuvent être pleinement traitées sur le plan plus détaillé de la *conscience* du lecteur. C'est donc directement et seulement en termes de la conscience, de ses structures instantanées et de la

Mentaux (§1.3.2).

dynamique de ces structures que l'on espère atteindre une compréhension plus raffinée des complexités de la fiction. *A priori*, les États Mentaux *ne figurent pas* dans la *réponse* à la question de la fiction. Pourtant, il *n'est nullement l'intention ici* de suggérer une quelconque inutilité *générale* à leur égard : rejeter les États Mentaux pour les propos de la question particulière posée ici n'implique aucune volonté *réformiste* relative à eux. Au contraire, la présente section est consacrée à montrer comment retrouver, en général, des États Mentaux dans le cadre proposé. Ce travail apportera plusieurs avantages. Non seulement le fait de représenter les États Mentaux dans un cadre qui, en théorie, aurait pris en charge pleinement le problème de la dynamique du mental pourrait avoir des conséquences heureuses sur la compréhension des États Mentaux et de *leur* dynamique. Mais de plus, la discussion de cette section permettra, dans §3.3, de représenter dans ce cadre les États Mentaux proposés comme spécifiques à la fiction, leurs rapports avec les États Mentaux usuels, et ainsi de mieux comprendre les *raisons* pour leurs insuffisances et la difficulté de ces rapports.

La stratégie employée pour aborder les États Mentaux a été déjà évoquée. Dans les cas traditionnels, une *attribution* des États Mentaux se fait dans un *bloc d'attribution* qui comporte des comportements du sujet dans certaines situations (§1.3.2). La seule différence dans l'attribution proposée ici consiste en ceci que le bloc comporte non pas des comportements du sujet observés à la troisième personne, mais plutôt les structures instantanées de la conscience à des instants différents. Il s'agit d'une *attribution* des États Mentaux dans un *bloc d'attribution* qui comporte *des structures instantanées de la conscience à des instants différents*.

Ce que l'on appelle ici bloc d'attribution a été rapproché à une théorie scientifique, dont les termes théoriques sont les États Mentaux⁶⁵. Ce rapprochement est fort utile ; en particulier, il permet de cerner un rapport entre les *attitudes* envers les termes théoriques ou les États Mentaux. À savoir, de même que l'on pourrait adopter des positions différentes par rapport aux entités théoriques posées par les théories scientifiques (réaliste, instrumentaliste, fictionnaliste *etc.*), de même on pourrait adopter des positions différentes relatives aux États Mentaux attribués au sujet (réaliste, instrumentaliste, anti-réaliste *etc.*). De même que les théories scientifiques peuvent se formuler et se développer sans se mêler dans les questions relatives au statut métaphysique de leurs entités, de même les théories qui attribuent des États Mentaux peuvent être développées sans prendre partie dans les débats relatifs à leurs statuts. En vérité, le scientifique

⁶⁵Ce genre de rapprochement se trouve chez Sellars (1997) et Lewis (1972) ; voir également §1.3.2.

souvent adopte cette attitude *neutre*, en délaissant la question métaphysique des entités théoriques de leurs théories. Ici, on adopte la *même attitude* envers les États Mentaux : des questions à l'égard de leur statut métaphysique ou ontologique sont totalement *hors sujet* ici. En particulier, il n'est question de prendre aucune position réaliste ni instrumentaliste ni anti-réaliste ni autre relative à « l'existence » des États Mentaux ; on est également neutre sur la question de savoir si les États Mentaux *causent* les actions auxquels ils sont en rapport ou *sont* des dispositions qui s'actualisent dans ces actions ou *sont* des raisons pour ces actions et ainsi de suite. À cette neutralité correspondra un relâchement de terminologie : on parlera parfois des croyances par exemple comme des « tendances » ou des « dispositions », on parlera de leurs « effets » et de leur « attribution ». Il ne s'agit que des aspects du bloc dans lequel les croyances sont attribuées. L'emploi du vocabulaire hybride sert seulement à étayer l'intuition et à alléger la discussion ; tous les termes sont employés, si l'on peut dire, dans leurs acceptions ordinaires, quotidiennes et *préthéoriques*, où ces distinctions et ces débats ne sont pas reconnus et n'ont que très peu d'importance. Le point mérite d'être répéter : *aucune position n'est adoptée ni soutenue à l'égard du statut des États Mentaux*.⁶⁶

La présente section s'organise autour du clivage entre les États Mentaux *narratifs* et les États Mentaux *non narratifs*. Dans §1.3.3, on a qualifié un État Mental de narratif si la base sur laquelle cet État est attribuée *doit* embrasser plusieurs instants, alors qu'un État Mental est dit non narratif s'il *peut* être attribué sur une base qui ne fait référence qu'à un seul instant. Alors que cette distinction est minorée dans un cadre strictement externaliste (§1.3.3), elle est significative dans le cadre « internaliste » adopté ici, dans la mesure où la structure instantanée de la conscience à un instant particulier pourrait permettre des attributions des États Mentaux non narratifs à cet instant. Il convient de souligner que cette distinction fait du sens même tenu compte du rapport étroit entre une structure instantanée et les structures instantanées à des instants voisins, thématiqué sous le nom de la continuité dans §2.3. Car ce dernier rapport a trait à l'interprétation des structures instantanées de la conscience (§2.4), et donc appartient au stade de la constitution des données (la succession de structures instantanées munie d'une interprétation), alors que la question de la « narrativité » des États Mentaux se pose ultérieurement à la mise en forme des données. La vitesse instantanée en physique réfère à un instant seul, la vitesse moyenne

⁶⁶La neutralité de ce travail par rapport à des positions et à des débats métaphysiques, qu'ils aient trait aux États Mentaux tels les croyances ou à la conscience, a été déjà affirmée à plusieurs reprises (par exemple, dans l'Introduction, dans §1.1.2, et dans §2.1.2).

a trait à plusieurs instants : ceci est indépendant du fait que, pour mesurer les vitesses instantanées d'un mobile, il faille mettre en jeu des observations relatifs à des instants divers.

Pour illustrer cette distinction dans le cadre proposé, on peut reprendre les exemples rencontrés dans §1.3.4 : comme État Mental non narratif, il y a le sentiment – une expérience phénoménale immédiatement présente (*occurrent*) – lequel s'oppose à l'émotion, État Mental narratif ayant trait à une gamme d'actions, de dits, de pensées, de sentiments, de tendances, relatifs à un intervalle temporel étendu⁶⁷. Dans le cadre « internaliste », l'attribution d'une émotion ne peut se faire que sur une base comprenant des états instantanés à *plusieurs instants*, où les différents pensées, sentiments, actions, événements (en tant qu'ils sont « observés » ou « éprouvés » à la première personne) sont en jeu ; par contraste, on peut apparemment attribuer un sentiment sur la base de l'état de la conscience à un instant seul. Or, il y a apparemment une distinction semblable dans le domaine cognitif : à l'opposé des croyances, qui sont des États Mentaux narratifs (§1.3.3), il semble y avoir un aspect cognitif « présent immédiatement à l'esprit » (*occurrent*). Lorsque le sujet tient compte de la pluie, la pluie est « présente à l'esprit », et si, à cet instant, il songe aux journées solaires, le soleil est lui aussi « présent à l'esprit » ; or seulement dans l'un des cas a-t-on trait à une croyance, comme l'atteste les états ultérieurs et antérieurs du sujet (où il décide de prendre un parapluie par exemple). Ou, pour prendre un autre exemple, un sujet pourrait avoir « à l'esprit » à un moment le fait que Gladstone était premier ministre, aussi bien que le « fait » que Sherlock Holmes est détective ; or seulement l'une de ces « cognitions immédiates » se rapporte à une croyance authentique, comme l'attestent ses actions, d'autres événements, et les états de la conscience à d'autres moments (il affirme que Holmes est fictionnel, par exemple). On appelle, en se permettant un éventuel léger éloignement par rapport au sens ordinaire, ces aspects cognitifs non narratifs, relatifs à un instant seul de la vie du sujet, des *pensées* : la pensée, au sens *technique* employé ici, est un terme générique pour toute « attitude » plus ou moins cognitive, envers quoi que ce soit (proposition, chose, événement, *etc.*), qui a trait à un instant particulier.

En résumé, il y a d'un côté les pensées et les sentiments, lesquels sont non narratifs, et de l'autre côté les croyances et les émotions, lesquelles sont narratives. Il s'agit dans cette section de rendre compte, dans le cadre proposé,

⁶⁷Pour rappeler le mot de Goldie (2000, p13) : « une émotion est structurée en sorte qu'elle constitue une partie d'un récit – en gros, une suite des actions et des événements, des pensées et des sentiments – dans lequel l'émotion s'enfonce. »

de tous ces États Mentaux ; d'ailleurs, la problématique de la fiction posée, qui concerne aussi bien son côté cognitif qu'affectif (§1.1), appelle une considération et des États Mentaux affectifs et des États Mentaux cognitifs. Dans les deux premiers paragraphes, on repère des aspects des structures instantanées de la conscience auxquels pourraient se rattacher les pensées et les sentiments ayant trait à ces structures (et à ces instants), pour en venir, dans le dernier paragraphe de la section, à la question des propriétés des successions des structures instantanées de la conscience auxquels les croyances ou les émotions peuvent se rapporter. Cet ordre de présentation est logique : dans la mesure où l'attribution des croyances et des émotions dépendent des structures instantanées de la conscience à plusieurs instants (et bien évidemment, des rapports entre ces structures), et dans la mesure où l'on peut attribuer à chaque structure, pris individuellement, des pensées et (éventuellement) des sentiments, il semble que l'on puisse aborder les croyances et les émotions à partir des pensées et des sentiments (successifs) du sujet.

2.5.1 Les pensées

Où il s'agit d'introduire et de considérer brièvement la pensée, conçue comme un État Mental cognitif « immédiat », et donc non narratif. Comme telle, une pensée à un moment donnée est associée à un aspect de la structure instantanée de la conscience, et hérite les propriétés de cet aspect. De même que le sujet pourrait ne pas tenir compte du fait d'avoir un aspect dans son état de la conscience à un moment, il peut ne pas tenir compte de certaines de ses pensées. Pour qu'il pense qu'il a une certaine pensée, il faut que le fait d'avoir cette pensée soit explicite dans sa conscience, il faut qu'elle soit focalisée.

Il s'agit dans cette section d'introduire l'État Mental que l'on appelle pensée, et de le rapporter au cadre proposé, c'est-à-dire aux structures instantanées de la conscience. Les pensées sont des États Mentaux cognitifs « immédiats à l'esprit ». Elles ont trait donc aux aspects de l'état instantané de la conscience à un instant donné. En effet, on montre, par le moyen de quelques exemples, que, conforme à la distinction entre les aspects focalisés et les aspects non focalisés de l'état instantané (§2.2.2), on peut distinguer entre des pensées selon que les aspects auxquelles elles ont trait sont focalisées ou non. Il est jusque-là question des pensées attribuées et conçues par un théoricien (fictif) qui observe la conscience du sujet (ou, comme on l'a dit dans §2.1.2, attribuées dans un point de vue à la troisième personne sur un point de vue à la première personne) ; pour

finir ce paragraphe, on considérera brièvement les attributions éventuelles par le sujet des pensées à lui-même.

En tant qu'État Mental « immédiat » (*occurrent*), une pensée individuelle a trait à un instant seul. Pour comprendre cette pensée dans le cadre proposé, il s'agit de la relier à la structure instantanée de la conscience à cet instant. En effet, une pensée pourrait être rattachée ou se référer tantôt à un objet ou à une configuration des objets de la structure instantanée, tantôt à une ou des relations de cette structure. Dans le premier cas, d'après les considérations de §2.2.2, on peut dire que la pensée est *focalisée*, dans le dernier cas, elle est plutôt *non focalisée*⁶⁸. Pour illustrer, on peut reprendre l'exemple du lecteur qui regarde sa bibliothèque et constate que deux livres sont adjacents (§2.2.2) ; on lui attribue la pensée que les deux livres sont adjacents. Il y a, dans la structure instantanée de la conscience à cet instant, trois objets – les deux *livres* et l'*adjacence* – reliées par une relation de « ... et ... entrent dans la relation de ... » (§2.2.2) ; la pensée que les deux livres sont adjacents est « au front » de la conscience à ce moment – elle est pleinement *focalisée*⁶⁹.

À l'opposé, il y a des pensées *non focalisées* (mais néanmoins présentes) qui, plutôt de se rapporter à des objets, sont associées à des relations de la structure instantanée de la conscience. À un instant où le sujet, ayant décidé de sortir, regarde vers sa fenêtre (il pleut) et puis décide de prendre un parapluie, la structure instantanée de sa conscience comporte les deux objets *il pleut*, et *prendre un parapluie*, reliés par une relation qui semble intuitivement être une sorte de conséquence. Comme soulignée dans §2.4, on ne peut qu'attribuer des États Mentaux à un agent que si l'on a interprété les aspects de ses structures instantanées. Lorsqu'on en vient à interpréter cette relation entre l'observation de la pluie (l'objet *il pleut*) et la décision de prendre un parapluie (l'objet *prendre un parapluie*⁷⁰), on la ramène à un certain nombre de faits et de préférences, dont par exemple le fait qu'un parapluie protège de la pluie, ou une préférence de ne pas se mouiller. En d'autres termes, on *associe* à cette relation, et par conséquent à ce moment, la pensée que le parapluie protège de

⁶⁸Pour des remarques sur la notion de *focalisation*, et de son rapport à la tradition, voir §§2.2.2 et 2.3.1.

⁶⁹Comme on l'a discuté dans §2.4, il est également possible d'interpréter un des objets de la structure par l'énoncé *les livres sont adjacents*. Les différences entre les deux interprétations ne sont pas profondes, mais elles touchent à l'expression de l'aspect auquel réfère la pensée : dans l'interprétation dans le texte, la pensée semblent se rattacher à la configuration des trois objets, alors que dans l'interprétation de §2.4, elle paraît avoir trait juste à l'objet *les livres sont adjacents*. On fait l'économie de ses minuties ici.

⁷⁰Dans cette conscience-là, cet objet peut être caractérisé comme une intention. On ne s'attarde cependant pas ici sur ce point.

la pluie. Ceci est un exemple d'une pensée qui est attribuée à la conscience à un instant particulier, mais qui ne correspond pas facilement à un objet ou même complètement à une relation. La pensée est plutôt « recélée » dans la relation entre l'observation et la décision ; elle est donc *non focalisée*. En outre, ceci est un exemple qui a le mérite supplémentaire de mettre en évidence la dépendance de l'attribution d'une pensée, et d'ailleurs des États Mentaux en général, à l'égard de l'interprétation des objets et des relations de la structure instantanée de la conscience. D'une manière générale, le plus complexe l'interprétation d'un aspect (la relation entre observation et décision ou prédiction), les plus complexes et détaillées les pensées qui sont attribuées à cet instant (dans cette interprétation).

Par ailleurs, il importe de souligner qu'il ne s'agit ici que des pensées, qui sont attribuées sur la base des instants individuels, et donc qui font référence à des structures instantanées individuelles, et nullement des croyances, qui réfèrent à plusieurs moments et donc à plusieurs structures instantanées. Donc, si l'on sait interpréter la relation comme ayant trait au fait que le parapluie protège de la pluie, et donc attribuer au sujet la pensée que le parapluie protège de la pluie à ce moment, on ne saurait toujours pas lui attribuer la croyance que le parapluie protège de la pluie sur la base de cette structure, avec cette interprétation, à ce moment. Bien entendu, cette relation *pourrait* être associée à une croyance, telle une croyance qui est en jeu lorsque le sujet décide de prendre son parapluie ; pourtant, pour parvenir à cette conclusion, il faut examiner d'autres structures instantanées à d'autres moments, pour voir par exemple s'il prend vraiment le parapluie, s'il agit de la même façon dans d'autres situations semblables, et ainsi de suite. On s'attardera sur ces questions dans §2.5.3, où l'analyse précédente et le concept de pensée s'avéreront utiles pour comprendre les croyances.

Jusqu'à présent, les aspects de la structure instantanée de la conscience discutés sont ce qu'un *observateur* (tels les théoriciens fictifs de §2.1.2) repérerait comme les pensées du sujet. Il *n'est pas nécessaire* que le sujet lui-même « soit conscient » de ces pensées : il pourrait bien avoir à un moment la pensée focalisée que les livres sont adjacents sans penser qu'il le pense. On termine ce paragraphe avec quelques remarques sur les cas où il tient compte de sa propre pensée, en l'occurrence du fait qu'il pense que les livres sont adjacents. Ce genre de cas peut être traité avec les concepts introduits précédemment (§2.2.2) : en effet, dans ce genre de cas, le *fait* d'avoir cette pensée est *focalisé*. Puisqu'une pensée focalisée correspond à un objet de la structure instantanée, il doit y avoir *non seulement* un objet qui « correspond » au fait que les deux livres soient adjacents, mais *de plus* quelque aspect structurel qui « caractérise » le

fait de reconnaître ceci comme une pensée⁷¹. Pour prendre juste une possibilité⁷², il peut y avoir un objet *mes pensées*, relié à d'autres objets qu'il « se reconnaît penser ». Cet objet pourrait apparemment être en relation aussi bien avec des objets interprétés comme énoncés, tels *les livres sont adjacents* qu'avec des objets autrement interprétés, tel un *livre* : dans le premier cas, il pense qu'il pense *que* les livres sont adjacents, dans le dernier cas, il pense qu'il pense *au* livre⁷³.

Ces considérations suffisent pour donner une idée des aspects de la structure instantanée de la conscience auxquels les pensées se rapportent, ou, si l'on préfère, des aspects de la structure qui « sont » les pensées. La question des sentiments est cependant plus complexe, dans la mesure où il n'est pas apparemment possible de repérer un objet particulier, une relation particulière ou une configuration particulière d'objets qui peuvent être rapprochées des sentiments.

2.5.2 Les sentiments

Où l'on soutient que le sentiment éprouvé à un moment peut être associé à une propriété globale de la structure instantanée de la conscience à ce moment. Ce faisant, on peut rendre compte de deux constatations à l'égard du sentiment – qu'il n'est pas réductible à une configuration particulière de pensées, et que l'on peut éprouver un sentiment sans tenir compte que l'on l'éprouve – dont l'une ou l'autre échappe à d'autres thèses relatives au sentiment.

Les sentiments sont des expériences phénoménales immédiates relatifs à des instants, alors que les émotions sont des États Mentaux narratifs qui ont trait

⁷¹ Cf. l'exemple de l'objet dans la structure où le sujet fait attention à l'adjacence des livres, §2.2.2. Voir aussi dans cette même section les réserves et les difficultés relatives à la « correspondance » entre un objet et une relation.

⁷² Voir, sur la possibilité de multiples interprétations d'une structure instantanée, (§2.4).

⁷³ À l'extrême de perspicacité, l'objet *mes pensées* serait en relation avec *tous les objets* dans la structure ; or, en pratique, ce n'est pas le cas. Par ailleurs, il est une question intéressante de savoir si l'objet *mes pensées* pourrait « correspondre » à quelque chose de non focalisé : car, s'il est conçu comme « correspondant » à une relation (d'arité un, donc, en d'autres termes, un prédicat) « ... sont parmi mes pensées », alors cette relation s'applique *en droit* à tout objet de la structure. D'où, en rapprochant les objets des visées des « vécus phénoménologiques », la thèse sartrienne que toute conscience est conscience « irréfléchie » et « non positionnelle » d'elle-même (Sartre, 1943, Introduction, §III). Pourtant, un prédicat qui s'applique à tout objet est un prédicat « vide », dans la mesure où il n'a pas d'effet sur la dynamique de la conscience et où l'état instantané et sa dynamique sont dans un rapport « symbiotique » (§2.3, notamment §2.3.4). Ainsi même si l'on l'admet, on ne saurait pas tirer des conclusions fortes de cette thèse sartrienne. Voir également la discussion de la « conscience fictionnelle » dans §3.1.1.

éventuellement à plusieurs sentiments, pensées, croyances, désirs, actions relatifs à des instants différents (§1.3.4). Dans ce paragraphe, on s'occupe des sentiments, et plus précisément de la question suivante : à quoi, dans la structure instantanée de la conscience à un instant, correspond-t-il un sentiment qui est éprouvé à cet instant ? Souvent, dans la littérature sur le sujet, la distinction admise ici entre sentiment et émotion n'est pas thématisée, de sorte que la question du rapport entre le sentiment et l'état instantané de la conscience est peu discutée, du moins sous ce nom. Néanmoins, on pourrait s'appuyer sur certaines discussions des *émotions*, en repérant certaines thèses à l'égard des émotions et en discutant *leurs analogues* sur le plan des *sentiments*. Précisément : on parviendra à la conclusion que les sentiments correspondent aux aspects « globaux » de la structure instantanée. Cette conclusion sera une sorte de synthèse de deux positions dont on trouve des versions dans la tradition : selon l'une, un sentiment est une certaine configuration de pensées (c'est-à-dire, des objets et des relations de la structure), alors que, selon l'autre, un sentiment est quelque chose de *sui generis* (un objet *sui generis* de la structure). Chacune de ces positions se veut fidèle à une des constatations suivantes relatives aux sentiments, en faisant fi de l'autre : d'une part, on peut éprouver un sentiment sans tenir compte (de manière focalisée) que l'on l'éprouve ; d'autre part, on ne sait pas associer à un sentiment une configuration particulière de pensées (d'objets et de relations de la structure) qui le résume. On présente d'abord ces deux constatations, ensuite les deux positions relatives aux sentiments, pour en venir enfin à la position qui est admise ici.

D'abord, de même que l'on pourrait penser au fait que deux livres sont adjacents sans que le fait d'avoir cette pensée soit lui-même focalisé, de même on pourrait éprouver un sentiment sans tenir compte de manière focalisé qu'on l'éprouve. Comme le dit Sartre « la peur n'est pas originellement conscience d'avoir peur, pas plus que la perception de ce livre n'est conscience de percevoir le livre »⁷⁴. En termes des objets, des relations et des structures instantanées de la conscience, cela veut dire qu'il y a un sentiment « dans » une structure instantanée de la conscience sans qu'il y soit nécessairement un objet focalisé qui « nomme » ou « réfère à » ce sentiment. (Car, rappelle-t-on, un objet correspondant au sentiment impliquerait que le sentiment soit focalisé ; voir §2.2.2.)

Ensuite, à la différence des pensées, on ne saurait pas repérer quelque objet, relation ou configuration d'objets qui « résume » exactement ou qui « correspond » à ou qui « est » le sentiment. On a vu dans le paragraphe précédent que les objets, les configurations des objets et les relations ont trait à des pensées :

⁷⁴Sartre (1995, p70).

or, il n'est pas certain qu'un sentiment est réductible à un certain nombre ou une certaine configuration de pensées. La peur d'un serpent ne consiste pas dans la pensée à ce serpent, ni dans la combinaison de cette pensée avec la pensée qu'il est dangereux (par exemple). Corrélativement la peur d'un serpent à un instant donné ne pourrait pas être « résumé totalement » par ou « identifiée » avec ces objets (*serpent*, *danger*) et ces relations dans la structure instantanée de la conscience.

Deux positions à l'égard du sentiment, qui peuvent être mises en rapport avec des positions traditionnelles et opposantes dans la littérature sur l'émotion, sont chacune fidèle à l'une de ces constations, en faisant fi de l'autre.

D'un côté, il y a la thèse selon laquelle un sentiment n'est qu'une certaine configuration de pensées ; dans des termes introduits dans §2.1.3, un sentiment est une certaine configuration des objets et des relations de la structure instantanée de la conscience. Cette thèse à l'égard des *sentiments* rappelle des thèses dites « cognitivistes » à l'égard des *émotions*, selon lesquelles une émotion n'est qu'une certaine configuration de croyances, de désirs et d'autres attitudes non affectives. Davidson, par exemple, tend à réduire les émotions à leurs causes et leurs raisons, lesquelles sont effectivement des attitudes, des croyances et des désirs⁷⁵. Ce genre de thèse fait fi de la difficulté apparente de spécifier exactement les pensées ou les objets et les relations qui sont pertinentes.

De l'autre côté, il y a la thèse selon laquelle un sentiment est un aspect *sui generis* ; pour le dire en termes des objets, relations et structures, un sentiment est un objet *sui generis* de la structure instantanée, qui ne pourrait pas être identifié avec aucune autre pensée « ordinaire ». On pourrait rapprocher cette thèse à l'égard des *sentiments* aux thèses selon lesquelles les émotions sont *sui generis* par rapport à d'autres États Mentaux. James, par exemple, soutient la thèse selon laquelle une émotion *est* le sentiment des changements corporels⁷⁶, thèse qui semble impliquer une certaine indépendance de l'émotion à l'égard d'autres attitudes⁷⁷. Cette position fait fi du fait que l'on peut éprouver un sentiment sans tenir compte de manière focalisée du fait que l'on l'éprouve,

⁷⁵Par exemple, il soutient (Davidson, 1976, p284) que la structure basique d'orgueil est que « la cause consiste premièrement dans une croyance relative à soi-même, qu'on manifeste un certain caractère, et deuxièmement dans une attitude d'approbation ou d'estime pour quelqu'un qui manifeste ce caractère. » Voir Solomon (1973) pour un autre exemple.

⁷⁶James (1884, p189-190).

⁷⁷Il s'agit ici d'une caricature de la position jamesienne, de même d'ailleurs que les autres positions présentées ici. *Stricto sensu*, il peut soutenir un rapport avec d'autres attitudes en tant que médiatisé par le corps, dans la mesure où les changements corporels qu'il associe au sentiment entrent en rapport avec la perception (*op cit.*, p189) et par suite avec les attitudes ayant trait à cette perception.

car si le sentiment est un objet particulier de la structure instantanée de la conscience, alors il est focalisé. Autrement dit, si le sentiment est un aspect *sui generis* de l'état instantané de la conscience, si, pour ainsi dire, la peur du serpent *s'étale à côté* du serpent « dans » la conscience du sujet, alors il est focalisé, contrairement à la constatation précédente.

Ces deux positions semblent trouver une sorte de synthèse dans une troisième position, selon laquelle le sentiment correspond à un « aspect global » de la structure instantanée de la conscience. Cette thèse à l'égard des sentiments se rapproche aux thèses à l'égard de l'émotion développées dans la tradition phénoménologique. Que l'on considère le mot de Sartre, selon lequel « l'émotion est une certaine manière d'appréhender le monde »⁷⁸. Le « monde » de quelqu'un qui a peur est concentré sur, et constitué presque exclusivement de, ce dont il a peur. Par ailleurs, cette idée que l'affect modifie « le monde » se trouve également chez Wittgenstein (au moins, selon une certaine lecture), lorsqu'il dit que « le monde de l'homme heureux est un autre monde que celui du malheureux »⁷⁹. Le pont entre le « monde » affecté et la structure instantanée de la conscience est immédiat : dans la perspective solipsiste adoptée ici, le « monde » à un instant particulier ne consiste qu'en ce qui est, d'une manière ou d'une autre, dans la structure instantanée de la conscience à cet instant⁸⁰. Le sentiment correspond donc, selon ce genre de thèse, à une propriété « d'ensemble » de la structure instantanée de la conscience. C'est-à-dire une propriété de la structure prise comme entier, une propriété « globale » de la structure, et non une propriété de certains de ces objets et de ces relations seulement. La structure instantanée de la conscience de celui qui a peur (à un moment lorsqu'il éprouve sa peur) est déséquilibrée en faveur de l'objet de sa peur, elle se restreint presque exclusivement à lui ; la *forme* de la structure instantanée de la conscience répond à son sentiment de peur.

Il s'avère que cette position rend compte des deux aspects constatés plus haut relatifs au sentiment, à savoir la difficulté en spécifiant des objets et des relations qui sont à associer au sentiment, et la différence entre éprouver et tenir compte que l'on éprouve un sentiment. En premier lieu, puisque le sentiment est éventuellement une propriété globale de la structure instantanée de la conscience, il se pourrait que tous les objets et toutes les relations de la structure lui soient pertinents. Or, puisque cette structure est ouverte, en ce sens qu'il est impossible de faire la liste de ses objets et de ses relations (§2.3.1), il n'est pas non plus nécessairement possible de faire la liste des objets et des relations

⁷⁸Sartre (1995, p71). Son terme « émotion » s'apparente à ce qu'on appelle ici « sentiment ».

⁷⁹Wittgenstein (1922, §6.43).

⁸⁰§1.1.2 et Wittgenstein, *op cit.*, §5.62 *sq.*

pertinentes au sentiment, d'où la difficulté en spécifiant quels objets et quelles relations se rapportent au sentiment. Les partisans de la position qui consiste à réduire les sentiments aux objets et relations de la structure instantanée de la conscience ont raison dans la mesure où le sentiment n'implique pas quelque chose en plus des objets et des relations ; ils ont cependant tort de supposer qu'il soit possible de fournir la liste complète des objets et des relations pertinents. En second lieu, si le sentiment renvoie à une propriété globale de la structure, à sa forme si l'on peut dire, alors il n'y a pas nécessairement d'objet dans la structure qui « correspond à » ou qui « réfère à » cette forme, et le fait que le sujet éprouve ce sentiment n'est pas nécessairement focalisé. En revanche, pour que ce fait soit focalisé, il faut qu'il y ait un objet dans la structure de la conscience qui lui « correspond » (§2.2.2), et cet objet devrait « correspondre » ou « référer » au sentiment, dans la même façon que l'objet adjacence « correspond » à la relation « à côté de » dans l'exemple de §2.2.2. Il y a donc bien une différence entre éprouver un sentiment et tenir compte que l'on l'éprouve. En outre, dans la mesure où le sentiment est associé à une propriété éventuellement globale de la structure, l'objet qui y « réfère » devrait « correspondre à » une propriété globale, plutôt qu'un objet, une relation ou une configuration d'objets particuliers ; en ce sens, cet objet est *sui generis*. Donc, la position qui soutient qu'un sentiment est un objet *sui generis* a raison dans la mesure où l'objet dans la structure qui « correspondrait au » sentiment est *sui generis* ; elle a cependant tort dans la mesure où il ne faut pas que cet objet soit présent dans la structure pour qu'un sentiment soit éprouvé – en vérité, il est seulement présent si le fait d'éprouver le sentiment est focalisé.

Dans cette thèse, on suit la tradition phénoménologique et la suggestion wittgensteinienne en posant qu'un sentiment est ou correspond à une propriété de la structure instantanée de la conscience prise comme entière. C'est-à-dire, généralement, que le sentiment est associé à une propriété globale, qui réfère à un nombre indéfini et indéfinissable des objets et des relations de la structure. Autrement dit, une propriété non pas d'un petit fragment de la structure, comme les objets et les relations pertinentes pour des pensées, mais une propriété de la structure en tant que totalité. On ne peut pas mettre suffisamment l'accent sur l'importance de la totalité de la structure. On a déjà souligné dans §2.3.4 l'importance de cette totalité pour la dynamique des structures instantanées, dans la mesure où, grâce au chevauchement des structures instantanées successives dans une dynamique conçue comme continue, elle contraint les structures instantanées qui peuvent suivre et précéder. Ceci n'est autre qu'une expression de l'éventuelle importance des sentiments pour la dynamique de la conscience.

2.5.3 États Mentaux narratifs

Où l'on voit que les États Mentaux narratifs, tels les croyances, peuvent être associés à des propriétés des successions des états instantanés de la conscience. De plus, en raison de la continuité de la dynamique de la conscience, ces propriétés se rapportent souvent à des aspects internes aux états instantanés à certains moments ; ainsi, les États Mentaux narratifs peuvent souvent être étudiés en examinant des états instantanés de la conscience à des moments particuliers. En conséquence, les propriétés de l'état instantané de la conscience mises en évidence dans §2.2 permettent une analyse raffinée de quelques phénomènes relatifs aux États Mentaux narratifs, dont certains qui se sont avérés difficiles pour les théories traditionnelles (§1.4). D'un autre côté, d'autres questions relatives aux États Mentaux narratifs, dont la question des leurs changements éventuels, se trouvent renvoyées à la dynamique des états instantanés de la conscience. Pour terminer, en guise d'éclaircissement du rapport avec d'autres approches des États Mentaux narratifs, on contraste l'analyse approfondie des États Mentaux ordinaires que pourrait offrir le cadre développé ici avec le « raffinement » proposé, sous la forme des États Mentaux sophistiqués, par des théoriciens comme Ramsey (§1.3.2).

Dans ce paragraphe, il s'agit de considérer la question de l'attribution des États Mentaux narratifs, c'est-à-dire des États Mentaux qui sont (nécessairement) attribués sur une base comportant des événements (expériences, aspects *etc.*) parsemés dans le temps (§1.3.3). Par exemple, la croyance s'attribue ordinairement sur la base de certaines « suites d'événements » où le sujet est dans une situation et ensuite se comporte dans certaines manières ; de façon semblable, comme on l'a vu dans §1.3.4, les émotions sont elles aussi associées à des « suites », comprenant aussi bien des actions et des événements que des sentiments. Dans ce paragraphe, on travaille seulement sur le cas de la croyance ; néanmoins, les éléments généraux que l'on repère s'appliquent *mutatis mutandis* à d'autres États Mentaux, à cette différence près que, là où seulement les pensées sont pertinentes aux croyances, les sentiments sont également pertinents pour certains autres États Mentaux telles les émotions. Naturellement, étant donné le travail du paragraphe précédent, la mise en jeu des sentiments ne pose pas des problèmes supplémentaires.

On adopte ici, comme l'a été souligné précédemment, une approche non révisionniste de la croyance, en s'intéressant pour l'essentiel à la conception

ordinaire et non rigoureuse de croyance (§1.3.2). Selon la tradition récente, les États Mentaux sont attribués à un sujet sur la base de ses comportements dans les situations où il se trouve (ou, on dit parfois, ses comportement et les « stimuli » qui l’affectent), observés à la troisième personne. Or, on ne dispose ici que des données à la première personne, c’est-à-dire les états instantanés de la conscience du sujet. Pour attribuer des États Mentaux sur cette base, il faut constater que la majorité des comportements et des situations pertinents pour l’attribution des États Mentaux sont observés par le sujet lui-même, de telle sorte que l’on pourrait appliquer les *mêmes* procédés d’attribution d’États Mentaux aux comportements et situations *en tant qu’il les saisit* (§1.5.4). On obtient ainsi les États Mentaux qu’il « aurait pu » attribuer à lui-même. Dans la mesure où les comportements et les situations en tant qu’il les saisit apparaissent de façon plus ou moins complexe dans ses états de la conscience à des instants particuliers, on arrive effectivement à attribuer des États Mentaux sur la base de ses états instantanés de la conscience⁸¹.

Plus précisément, une croyance particulière est attribuée sur la base des *successions des états instantanés de la conscience*. Dans un premier temps, on considère, sans entrer dans les détails, les genres d’aspects de ces successions que l’on associe normalement aux croyances ordinaires, afin de montrer la possibilité d’attribuer des croyances ordinaires dans le cadre proposé. Or, le fait de retrouver les États Mentaux narratifs dans ce cadre permet de se servir des outils conceptuels développés dans §§2.1–2.3. En particulier, en conséquence de la continuité de la dynamique de la conscience (§2.3.4), une croyance, qui est d’abord associée aux aspects des successions des états instantanés, pourrait être souvent pensée en termes des aspects de l’état de la conscience *à un instant particulier*. Ce qui permet parfois d’aborder les croyances *dans* des états instantanés de la consciences à des instants particuliers ; par conséquent, les propriétés des états instantanés mises en évidence dans §2.2 s’avèrent utiles pour comprendre les croyances, offrant d’ailleurs la possibilité de rendre compte de certaines des aspects des croyances relevés dans §1.4.2 comme difficiles pour les théories traditionnelles. D’un autre côté, après avoir ramené des croyances aux aspects des successions des états instantanés, d’autres questions à leur égard, dont les questions relatives à leur « sens » et à leur changement, se présentent sous un nouveau jour : elles renvoient effectivement à la dynamique de la conscience. Jusque là, il ne s’agira que des États Mentaux ordinaires, lesquels sont employés habituel-

⁸¹ Comme on l’a déjà fait remarquer dans la note 132 de §1.5.4, la question de la conception appropriée du contenu des États Mentaux, et en particulier de sa dépendance sur des facteurs externes au sujet, a peu de rapport avec le but de la présente discussion. Pour des considérations supplémentaires, voir la note 59 de §2.4.

lement dans la vie quotidienne, et non pas des États Mentaux sophistiqués tels les degrés de croyance et préférences proposés par la théorie de la décision (voir §1.3.2), qui prétendent en un certain sens « raffiner » la conception ordinaire. Pour terminer cette section, on comparera l'analyse approfondie des croyances que l'on pourrait s'attendre d'une théorie mûre de la dynamique des états instantanés de la conscience avec le « raffinement » proposé par les partisans des États Mentaux sophistiqués et notamment de la théorie de la décision.

Croyances comme propriétés des successions des structures instantanées de la dynamique

Pour le dire de manière très approximative, les croyances ordinaires sont associées à des dispositions ou des tendances. Ceci est admise par les théoriciens des croyances sophistiquées aussi bien que par les théoriciens des croyances ordinaires⁸². Comme on l'a souligné au début de la section, la question de la teneur métaphysique ou autre de ce rapport entre croyance et disposition (la croyance est-elle une disposition ? ou correspond-t-elle à une disposition ? ou la disposition est-elle l'effet d'une croyance ?) est carrément *hors sujet* ici, si bien que l'on parle des croyances comme étant, correspondant à, étant associées à, ou même ayant pour effet des dispositions sans différence entendue entre ces expressions. Corrélativement, on dit qu'une croyance est « effective » ou « en jeu » ou qu'elle « s'exerce » lorsque la disposition est « en jeu » ou « s'exerce » (lorsqu'elle est, pour ainsi dire, sur la voie vers l'actualisation).

Comme l'a fait remarquer Ryle, le terme disposition ou tendance, lorsqu'il est associé à la croyance, ne signifie pas une disposition à faire une seule chose ou à se comporter d'une seule façon, mais une disposition plus souple, ouverte, et imprécise⁸³. Pour ainsi dire, dans les situations différentes où la disposition ou la croyance est *en jeu*, elle *s'actualise* ou *s'exerce* dans des manières différentes. D'une manière générale, on peut associer à des « conduites » du sujet des croyances ou des dispositions qui sont en jeu lors de ces « conduites » : souvent, la même croyance, ou la même disposition, intervient de manières différentes dans des « conduites » différentes relatives aux situations différentes. Dans la suite, pour considérer une croyance et son rapport aux états instantanés de la conscience, on considère *des « conduites » particulières* auxquelles cette

⁸²Voir Ryle (2000, p114) et Ramsey (2003, p164) (les citations sont reproduites dans la note 56). La différence remarque-t-on réside dans le sens rigoureux ou autre qu'on accorde à cette disposition ; voir §1.3.2.

⁸³« Les termes dispositionnels tels 'connaître', 'croire', ... signifient des capacités, des tendances ou des inclinations à faire, non pas des choses d'un seul et unique genre, mais des choses de plusieurs genres. » (Ryle, *ibid.*)

croyance peut (ou non) être pertinente.

Par exemple, on associe intuitivement la croyance que le parapluie protège de la pluie à une disposition à un certain nombre de comportements différents dans des circonstances différentes, tels que prendre le parapluie lorsqu'il risque de pleuvoir, le porter dans la pluie, affirmer cette croyance en cas où l'on est questionné là-dessus, et ainsi de suite. Dans chacune de ces circonstances *particulières*, la croyance ou la disposition est ou pourrait être en jeu, et certaines « conduites » correspondent à « l'actualisation de l'une des facettes de la disposition » ou « une exercice » de la croyance. Chacune de ces circonstances et conduites est, remarque-t-on, relativement *locale*, ayant trait à un intervalle temporel limité. Dans une circonstance où le sujet regarde par la fenêtre et voit la pluie, sa « conduite » ultérieure, consistant à prendre le parapluie en sortant, représente une « actualisation » de (une facette de) la disposition, un « effet » ou une « exercice » de la croyance ; évidemment cette circonstance et cette conduite n'occupent qu'un intervalle temporel limité.

Dans le cadre adopté ici, la conduite du sujet se rapporte à une certaine succession des structures instantanées de la conscience, laquelle mène d'une structure instantanée où le sujet remarque la pluie à une structure instantanée où il sort avec le parapluie. Ainsi, de la même façon que l'on associe la croyance que le parapluie protège de la pluie à la suite d'événements observés publiquement, on peut associer la croyance à cette succession de structures instantanées de la conscience. L'« exercice » de la croyance dans une circonstance particulière, sa « mise en jeu » dans cette circonstance, « l'actualisation » de la disposition qu'on lui associe dans cette circonstance, a trait à ou réfère à ou est une propriété de la succession des structures instantanées de la conscience, en l'occurrence, la propriété de mener des structures instantanées où le sujet remarque la pluie aux structures où il sort avec le parapluie. Lorsque l'on attribue des croyances du point de vue d'autrui, il s'agit des « successions » d'événements ; de même ici, il s'agit des successions des structures instantanées de la conscience.

Il est donc possible d'attribuer une croyance à un individu dans le cadre proposé, en regardant les successions des structures instantanées de la conscience, et en y repérant des propriétés qui ressemblent aux propriétés des suites des événements, observés à la troisième personne, que l'on rattache habituellement à la croyance. Tout ce que l'on fait ordinairement, on peut reproduire, *grosso modo*, dans le cadre proposé⁸⁴. En effet, on peut faire plus, dans la mesure ce

⁸⁴À des réserves près relatives à la question de l'« externalisme » en matière du contenu des États Mentaux, qui ont été déjà constaté dans la note 132 de §1.5.4. Voir également la note 59 de §2.4.

cadre procure des outils puissants pour analyser les successions des structures instantanées de la conscience.

La croyance *sur* et *dans* les structures instantanées de la conscience

Une croyance, lorsqu'elle « s'applique » ou « se rapporte » à la conscience, le fait d'abord donc sur le mode *non instantané*, en tant qu'aspect des successions des structures instantanées de la conscience. Pourtant : une représentation correcte des structures instantanées de la conscience, permettant de penser la dynamique de ces structures instantanées comme continue, *implique* que *certains aspects dynamiques des successions de structures instantanées apparaissent* d'une manière ou d'une autre *comme des aspects à l'intérieur des structures instantanées à des instants particuliers*. Ce constat, voire cette contrainte sur une théorie susceptible à supporter une pensée adéquate de la dynamique, était le thème de §2.3 et notamment de §2.3.4. À force de répéter un point maintes fois souligné là, le sens de la continuité de la dynamique réside dans le rapport étroit entre l'intérieur de l'état instantané et son extérieur, en l'occurrence, les états qui le succèdent et le précèdent dans la dynamique. Évidemment, dans la mesure où la continuité est une notion locale, ayant affaire aux états instantanés à des instants voisins, cette « réflexion » des aspects *de* la dynamique *dans* les états instantanés à des instants particuliers ne se fait que pour des aspects dynamiques locaux, c'est-à-dire relatifs aux successions courtes. Or, on vient de remarquer que les conduites et les circonstances particulières où les croyances se mettent en jeu sont elles mêmes *locales*, c'est-à-dire limitées dans le temps. De telle sorte que, dans la considération des conduites particulières du sujet, il est parfois possible de profiter de ce rapport entre la dynamique de la conscience et la structure instantanée de la conscience à un instant. De cette manière, on peut repérer quelques aspects *dans* les structures instantanées de la conscience à certains instants, qui « correspondent » aux propriétés *des successions* de ces structures instantanées, lesquelles, pour leur part, sont associées à des croyances particulières. En un mot, comme anticipé dans §2.3.4, on pourra parler des croyances en jeu, lesquelles sont d'abord associées à certaines *successions* de structures instantanées, *dans* les structures instantanées à des moments particuliers. Un exemple éclairera ce point.

Dans la circonstance discutée ci-dessus où le sujet regarde la pluie hors de sa fenêtre et puis sort avec son parapluie, il s'agit d'une succession des états instantanés de la conscience. Au début de cette succession, le sujet remarque la pluie

(la pluie est focalisée), laquelle correspond à un objet *il pleut*⁸⁵ dans la structure instantanée relative à cet instant (appelons-la S_1). À la fin de cette succession, au moment où il saisit le parapluie, son attention est au moins partiellement prêtée à ce parapluie, et notamment au fait de le prendre : dans la structure instantanée à cet instant (appelons-la S_2), il y a l'objet *prendre le parapluie*⁸⁶. On associe sa croyance que le parapluie le protège de la pluie au rapport dynamique, de succession, entre ces structures instantanées de la conscience (S_1 et S_2).

Pourtant, comme on l'a vu dans §2.3.3, dans une représentation *continue* de la dynamique, les structures instantanées de la conscience à des instants voisins *se chevauchent*, c'est-à-dire, ils ont des aspects en commun (ou, pour reprendre la terminologie plus prudente préconisée dans §§2.2.1 et 2.4, des aspects qui « correspondent » les uns aux autres). Dans le cas des deux structures instantanées aux bouts de la succession, S_1 et S_2 , l'une avec l'objet *il pleut*, l'autre avec l'objet *prendre le parapluie*, le chevauchement n'est pas évident. Mais le *rapport* entre les deux structures est clair : il tient au rapport entre les deux objets *il pleut* et *prendre le parapluie*, rapport que l'on associe intuitivement à une sorte de « passage de l'observation à l'action », une sorte de « prise de décision » sur la base d'une observation, une sorte de « d'inférence à l'action », pour ainsi dire. Normalement on dit que la croyance relative aux propriétés protectrices des parapluies étaye (ou justifie) cette « inférence ». Dans la mesure où les deux structures ne contiennent chacune que l'un seulement des objets pertinents, ni l'une ni l'autre de ces structures instantanées comporte ce rapport entre les deux objets comme *relation* de la structure. Néanmoins, dans la mesure où la succession de structures instantanées est continue, et dans la mesure où cette continuité est caractérisée par un chevauchement, il faut que l'une des structures instantanées à un instant intermédiaire chevauche avec les *deux* structures S_1 et S_2 ⁸⁷. Une telle structure comporte les deux objets *il pleut* et *prendre le parapluie* et une relation entre eux qui correspond au rapport « d'inférence à l'action »

⁸⁵On rappelle encore le sens large d'objet employé ici, §2.1.3.

⁸⁶On rappelle qu'e l'on discute ici des exemples typiques et tout probablement simplifiés (§§2.3.1 et 2.4).

⁸⁷Plus précisément : il faut que ces deux structures soient « liées » par une chaîne de structures instantanées, ordonnées par l'instant approprié, telle que chaque paire de structures adjacentes se chevauchent. On considère dans le texte l'exemple simple d'une seule structure intermédiaire. Remarquons que, puisque la structure est ouverte et la notion ordinaire d'instant temporel est peu raffinée (voir note 38 et §2.3.3), pour des conduites à courte durée, il est non seulement possible, mais effectivement nécessaire pour éviter une théorisation excessive et peu soutenue, de penser en termes d'une seule structure (ouverte) intermédiaire qui se chevauche avec les deux autres. Voir également la note 54.

discuté. Une telle structure chevauche la structure précédente (S_1) et la structure suivante (S_2) dans la mesure où elles ont des objets en communs. Elle fait le pont, pour ainsi dire, entre les deux structures S_1 et S_2 , pont qui est appelé par la continuité de la dynamique des structures instantanées. Enfin, et voici le point capital, cette structure *comporte une relation entre il pleut et prendre le parapluie* que l'on interprète comme « l'inférence à l'action » et *que l'on associe à la croyance que le parapluie protège de la pluie*. Dans ce sens, on identifie un aspect interne à cette structure instantanée à une croyance qui se pense d'abord comme propriété de la succession des structures instantanées dont cette structure fait partie. On pense, avec cette structure à cet instant, la croyance comme aspect d'une structure instantanée.

Ainsi, aux propriétés des (courtes) successions des structures instantanées de la conscience correspondent des aspects – objets, relations, configurations d'objets – *dans* certaines de ces structures. Mais le rapport entre l'état instantané et la dynamique qui est impliqué par la continuité est symétrique : non seulement la dynamique d'un système considéré comme continu pose des exigences sur les états instantanés (le chevauchement des états instantanés successifs dans le temps), mais à l'inverse un état instantané pose des contraintes sur la dynamique (seuls des états avec lesquels il chevauche pourrait lui suivre ou précéder immédiatement). On a déjà vu dans §2.3.4 comment les propriétés de la *totalité* de la structure, par exemple, peuvent avoir des conséquences importantes sur les structures qui la suivent immédiatement. Dans cette mesure, les aspects d'une structure instantanée de la conscience, dont par exemple les objets et les relations qui y sont importants ou « centraux », sont signifiants du point de vue de la dynamique de la succession dont la structure fait partie. Pour revenir à l'exemple, tant qu'une structure instantanée de la conscience comporte des objets *il pleut, prendre un parapluie* et une certaine relation entre eux, il est plus probable (bien que non certain, comme on discutera prochainement) que le sujet décidera de prendre le parapluie dans les structures instantanées de la conscience qui suivent.

Le rapport « symbiotique » entre l'état instantané et sa dynamique (§2.3) a donc des conséquences pour la question des croyances. Des propriétés d'une succession *locale* de structures instantanées qui sont associées à une croyance (ou une disposition) qui est « en jeu » ou « effective » pendant ces moments peuvent être également mises en correspondance avec les aspects internes à certaines structures instantanées à certains instants particuliers. À l'inverse, des aspects à l'intérieur d'une structure instantanée peuvent, par leur portée dy-

namique, correspondre à des tendances, et donc éventuellement des croyances, qui peuvent ou non s'actualiser effectivement dans les structures instantanées suivantes. Il y a donc une *corrélation* entre les *croyances* qui sont *en jeu* ou *effectives* à des moments particuliers et les *aspects* des *structures instantanées* à ces moments (et à des moments voisins). En relâchant un peu l'expression, on pourrait parler des croyances aussi bien *sur* que *dans* les structures instantanées de la conscience.

Conséquences

La corrélation entre les propriétés des successions des structures instantanées de la conscience et les aspects internes aux structures instantanées individuelles qui peuvent être associés aux croyances est une des conclusions principales du présent paragraphe. Il en découle et s'y accroche deux conséquences importantes.

Premièrement, la localité de la structure instantanée de la conscience (§2.2.1), et la notion de *hors sujet* qui va de pair devient un élément crucial dans la considération des croyances (et des États Mentaux en général). Le fait pour une croyance au sujet de X d'être « en jeu » dans un intervalle temporel entraîne ou correspond aux propriétés de la succession des structures instantanées de la conscience pendant cet intervalle, qui, par la corrélation, se rapportent aux aspects de certaines structures instantanées à certains instants pendant cet intervalle, aspects qui peuvent pour leur part être associés à cette croyance et donc à X . Donc, par contraposition, l'absence, relative aux structures instantanées dans un intervalle temporel, des aspects qui peuvent être associés à X indique que la croyance du sujet relative à X n'est pas en jeu à ces moments. Par exemple, aux moments où le sujet qui prend son parapluie, les structures instantanées de la conscience ne comportent aucun objet, relation ou aspect qui a trait à la croyance du sujet relative à la couleur des yeux d'un ancien camarade de classe ; corrélativement, cette croyance n'est pas « en jeu » ou « effective » pendant ces moments. Bref, le degré auquel une croyance est effective ou en jeu dans un intervalle temporel particulier est relié à son « degré de *présence* » dans certaines structures instantanées de la conscience à certains instants de cet intervalle.

Pour bien saisir ce point, il convient de rappeler un certain nombre de propriétés du cadre théorique qui ont été évoquées auparavant. Tout d'abord, on remet l'accent sur le vague des structures instantanées, figure omniprésente à toutes ces considérations (§2.3.1). Sans doute y a-t-il un rapport entre la « gradation vague » dans l'appartenance des aspects à des structures instantanées

et le degré auquel les croyances qui correspondent à ces aspects sont en jeu : moins un aspect appartient (certainement) à quelque structure instantanée dans un intervalle temporel, moins la croyance qui correspond à cet aspect sera « en jeu » ou « effective ». Par exemple, lorsque le sujet sort avec son parapluie, il se pourrait que la raison pour sortir soit à l'arrière-plan de ses structures instantanées de la conscience (par exemple comme objet dont l'appartenance à ces structures est douteuse) ; en ce cas, on peut dire que sa croyance à cet égard est faiblement « en jeu » ou faiblement « active ». Bien que ce ne soit sans doute pas le seul facteur qui détermine si une croyance est plus ou moins importante ou effective à un moment⁸⁸, l'appartenance plus ou moins « pleine » d'un aspect à une structure instantanée est importante à cet égard. Par ailleurs, il importe de rappeler que le caractère focalisé ou non d'un aspect est indépendant de son « degré d'appartenance » à la structure instantanée (§2.3.1). Donc, si l'on relie le « degré d'activité » d'une croyance à son « degré de présence » dans certaines structures instantanées, *il ne s'ensuit pas* que la croyance doit y être présente en tant qu'aspect focalisé. Et en effet, une croyance pourrait se rapporter aussi bien à un *objet* d'une structure qu'à une *relation*, ou, pour reprendre le terme de pensée, introduit et discuté dans §2.5.1, à une pensée *focalisée* à un certain instant, aussi bien qu'à une pensée *non focalisée*, tout en étant pleinement « en jeu » ou « active ». Dans l'exemple ci-dessus, la croyance que le parapluie protège de la pluie est associée à une *relation* entre deux objets dans une structure (autrement dit, à une pensée non focalisée), alors que la croyance qu'il pleut correspond à un objet d'une structure instantanée, à savoir l'objet *il pleut* (autrement dit, à une pensée focalisée). Or, il est évident de cet exemple que le caractère focalisé ou non n'a pas de conséquence sur le « degré d'activité » de la croyance : la croyance relative aux propriétés protectrices des parapluies est très « active » et importante, alors même qu'elle est non focalisée dans les structures instantanées à des instants appropriés.

Une deuxième conséquence de cette corrélation entre croyances, associées *de prime abord* aux propriétés des successions des structures instantanées de la conscience, et aspects internes de certaines structures à certains instants, est le rapport avec les faiblesses des États Mentaux ordinaires mises en évidence dans §1.4. On y a vu que les attributions ordinaires des États Mentaux ne permettent guère de comprendre le rapport, souvent complexe, entre les États Mentaux à des instants particuliers : on a même parlé des rapports *d'activité* entre États Mentaux qui étaient peu compris. En outre, il a semblé que cette

⁸⁸La forme de la structure, ou la manière dont les objets et les relations y sont organisées, en est apparemment un autre.

difficulté soit liée à une autre difficulté, à savoir celle du rapport entre les États Mentaux attribués et les expériences phénoménales du sujet, c'est-à-dire, ses états instantanés de la conscience. Dans le cadre proposé, ces difficultés ne se posent plus. Dans la mesure où les croyances qui sont « en jeu » ou « effectives » correspondent à des aspects présents dans des structures instantanées de la conscience à certains instants, on pourrait aborder ce rapport d'*activité* entre des États Mentaux en termes des rapports entre les aspects des structures instantanées de la conscience qui leur sont associés. Les deux exemples qui ont posé problème pour des conceptions traditionnelles des États Mentaux dans §1.4.2 attestent la fécondité de l'approche proposée ici. L'un concerne le rapport entre certaines croyances actives et pertinentes actuellement (celles relatives au passage du texte que vous lisez actuellement) et d'autres croyances qui ne sont pas « à l'esprit » (celles relatives à la couleur des yeux d'un ancien camarade de classe). Dans le cadre proposé, la différence d'« activité » entre ces croyances est immédiatement claire : les premières correspondent aux aspects *dans* la structure instantanée de la conscience à ces instants, les dernières y sont *absentes* ; elles sont *hors sujet*. L'autre exemple concerne les moments d'hésitation devant une décision, où l'on pèse les options. Dans ce cadre, les rapports entre les différents États Mentaux aux moments d'hésitation sont à comprendre en termes des rapports, éventuellement complexes et difficiles, entre les aspects dans les structures instantanées de la conscience à ces moments qui leur correspondent. Quant à la prise de décision finale, alors que, par la continuité, les structures instantanées pourraient apporter des indices, c'est largement une question pour la *dynamique* des structures instantanées de la conscience, question vers laquelle il faut maintenant se tourner.

L'importance de la dynamique

Bien entendu, la corrélation entre les propriétés des successions des structures instantanées de la conscience et les aspects des structures individuelles *n'est pas parfaite*. Notamment, alors qu'il y a ce rapport entre la structure instantanée à un instant et les structures qui la succèdent et la précèdent, celles-ci ne sont nullement déterminées par celle-là. En un mot, la question pleine de la succession des structures, et par conséquent des structures qui suivent et précèdent une structure particulière appartient à la dynamique. On n'a esquissé que quelques pas préliminaires vers une théorie pleine de la dynamique des structures instantanées de la conscience dans §2.3. Néanmoins, il convient de souligner l'importance de cette question en indiquant trois genres de questions importantes relatives aux croyances qui s'y ramènent.

Premièrement, pour une structure instantanée particulière, la question de savoir si un aspect de la structure correspond à une croyance ou seulement à une pensée relève de la dynamique. Ce point est évident : la pensée est entendue ici comme un Etat Mental qui s'attribue sur la base des structures instantanées individuelles, indépendamment des structures qui les succèdent ou qui les précèdent, alors que la croyance, comme on l'a vu, s'associe premièrement aux successions des structures plutôt qu'aux structures individuelles. Donc, comme on l'a remarqué dans §2.5.1, on peut certainement associer ou attribuer la pensée qu'il pleut à une structure instantanée qui comporte l'objet *il pleut*, alors que cette pensée ne correspond à une croyance que si la succession des structures instantanées de la conscience dont elle fait partie possède certaines propriétés. Si par exemple il s'agit d'un instant pendant une hallucination ou autrement pendant l'immersion fictionnelle, on ne peut pas dire que la pensée qu'il pleut correspond à une croyance. Ce point mérite d'être souligné afin d'éviter des confusions. Dans les discussions des croyances, on profite souvent de la possibilité de repérer des aspects des structures instantanées de la conscience à des instants particuliers qui peuvent être associés à telle croyance, pour aborder des questions relatives à cette croyance dans le cadre d'une structure instantanée particulière à un instant particulier (et bien choisi). Ce faisant, on discute à strictement parler des pensées à un certain instant, en *sous-entendant* que les structures instantanées qui suivent et qui précèdent sont telles que ces pensées sont vraiment des croyances. Il ne s'agit à *aucun moment* de supposer qu'une croyance est un aspect d'une structure instantanée seule. Ce point s'applique à toutes les discussions relatives à des États Mentaux narratifs, dont celles qui portent sur des émotions ou des *faire comme si*.

Deuxièmement, pour pouvoir identifier une pensée à un instant avec une croyance, *il n'est pas nécessaire* que la disposition associée à cette croyance s'actualise effectivement. Il se peut que, à un certain instant, la structure instantanée comporte les deux objets *il pleut* et *prendre un parapluie* avec la relation entre eux que l'on associe normalement avec la croyance que le parapluie protège de la pluie, alors que, au prochain instant, le sujet rappelle qu'il a perdu son parapluie (la structure à cet instant comporte un objet comme *parapluie perdu*). Dans un tel cas, le sujet sort alors sans parapluie, ce qui contredit la disposition de prendre un parapluie lorsqu'il pleut, disposition que l'on associe normalement à la croyance que le parapluie protège de la pluie. Bien que la disposition ne « s'actualise » pas en vérité, bien que la croyance ne « mène » pas au bout, pour ainsi dire, on ne renoncera pas à associer la relation dans la structure avec la croyance : il s'agit plutôt d'une raison prépondérante qui l'emporte sur la

disposition et qui « annule » l'effet « ordinaire » de la croyance.

Cette possibilité de présence d'une croyance sans son « effet » ordinaire, de pertinence de la disposition associée sans son actualisation, n'est nullement contradictoire. Au contraire, il appartient au sens du terme disposition de *ne pas s'actualiser* dans certaines situations où elle pourrait s'actualiser et où elle s'actualiserait « normalement ». Comme on l'a déjà remarqué (§1.3.2), dans le cas des États Mentaux ordinaires qui sont employés dans la vie quotidienne, la notion de disposition n'est pas rigoureuse. Dans le cadre proposé ici, où les dispositions associées aux croyances sont des propriétés des successions des structures instantanées, il appartient à la dynamique d'approfondir les détails touchant à une disposition. C'est-à-dire, la question des circonstances où une disposition s'actualise ou non, et par conséquent la question des aspects qui empêchent ou encouragent certains « effets ordinaires » des croyances, sont des questions pour la dynamique. Dans ce sens, si la notion ordinaire et non rigoureuse de disposition ou de croyance est à préciser, cela relève du moins partiellement de la compréhension de la dynamique de la conscience⁸⁹.

On pourrait venir, et c'est la troisième remarque, à une conclusion semblable à l'égard d'autres questions relatives aux croyances : elles renvoient également à la dynamique. Par exemple, la question du changement des croyances revient à une question sur les propriétés de plus en plus « globales » de la dynamique des structures instantanées : dans la mesure où une croyance en tant qu'elle est « effective » ou en « jeu » correspond à une propriété d'une succession courte des structures instantanées de la conscience, des changements des croyances ont affaire aux changements dans les propriétés des successions locales. Ou, autre exemple, les éventuelles variations du « sens » ou des « modalités » d'un État Mental pourrait être pensées à partir de la dynamique des structures instantanées de la conscience⁹⁰. Le « sens » ou la « modalité » de l'attribution de l'État Mental dans une situation particulière est associé à une propriété de la succession de structures instantanées impliquées dans cette situation, de telle sorte que les divergences et les convergences entre les « applications » ou les « sens » de l'État Mental dans des situations différentes – « l'unité » ou la « diversité » de son « sens » ou son « usage », pour ainsi dire – se livre dans la comparaison de ces propriétés des successions des structures.

Il ne s'agit pas de discuter ces aspects en profondeur, car une telle discussion

⁸⁹Étant donné le caractère linguistique des termes disposition ou croyance, l'autre partie est naturellement une étude de l'usage ordinaire de ces termes.

⁹⁰C'est le troisième point relevé contre les théories traditionnelles des États Mentaux dans §1.4.2 ; voir aussi Wittgenstein (1953) par exemple.

s'élignerait trop du sujet de cette thèse ; il s'agit seulement de souligner le rôle clef de la dynamique. Dans la mesure où les États Mentaux sont associés dans des manières plus ou moins complexes à des aspects plus ou moins complexes des successions de structures instantanées de la conscience, on pourrait dire qu'ils sont « émergent » depuis ces successions de structures instantanées, à la façon de la température d'un gaz qui est « émergente » depuis des propriétés de ses particules individuelles. Or, pour grande que soit l'utilité des États Mentaux en tant que concepts, on a suffisamment montré (§1.4) que des questions *détaillées* et *épineuses* à leur égard, à leurs changements, à leur sens précis et ses variations éventuelles, ne trouvent pas de réponse sur le niveau des États Mentaux eux-mêmes : si elles trouvent des réponses, ce serait sur le niveau inférieur – celle des structures instantanées de la conscience et leur dynamique.

Résumé : États Mentaux ordinaires

Dans les sections précédentes de ce chapitre, on a jeté les bases d'une théorie du mental qui prend son départ de la conscience. Cette théorie est d'ores et déjà suffisante pour aborder des questions relevées à l'égard de la fiction, comme on le verra dans §§3.1 et 3.2. Or, on a voulu montrer également que ce cadre n'entraîne aucune impossibilité de parler et d'employer les États Mentaux. Dans les deux paragraphes précédents, on a montré que les États Mentaux non narratifs, à savoir les pensées et les sentiments, se rapportent aux différents aspects de la structure instantanée de la conscience à l'instant approprié ; en l'occurrence, aux objets, configurations des objets ou relations dans le premier cas, aux propriétés globales dans le dernier cas. Le présent paragraphe est consacré aux États Mentaux narratifs, dont l'attribution réfère à plusieurs instants, et discute principalement l'exemple des croyances.

Il n'est pas difficile de voir à quoi correspond les croyances dans l'approche proposée : plutôt que propriétés de successions d'événements observés de l'extérieur, pour ainsi dire, il s'agit des propriétés correspondantes des successions des états instantanés de la conscience, où ces événements apparaissent en tant que saisi de l'intérieur, si l'on peut dire. En revanche, ce sont les conséquences de l'application des analyses effectuées dans les premières sections de ce chapitre (notamment §§2.2 et 2.3) aux successions pertinentes pour les croyances qui sont intéressantes. En effet, la continuité de la dynamique de la conscience a pour conséquence que les propriétés des successions des structures instantanées que l'on associe aux croyances peuvent être mises en rapport avec des aspects des structures particulières appartenant à ces successions. Il s'ensuit que l'on peut souvent aborder des questions relatives aux croyance en étudiant

les structures instantanées à des moments appropriés. On a vu, par exemple, que le fait pour un aspect d'être absent de toutes les structures instantanées d'une succession implique que l'éventuelle croyance du sujet ayant trait à cet aspect n'est pas *en jeu* à ces moments. Il paraît que certaines propriétés difficiles des croyances relevées dans §1.4, ayant trait à leur « degré d'activité » ou aux rapports complexes entre elles à des moments particuliers, peuvent être comprises de cette manière. Il reste que d'autres questions à l'égard, telles la question des conséquences finales de leur mise en jeu, ou la question de leur changements, doivent se penser à une échelle plus globale. Elles renvoient finalement à la dynamique des états instantanés de la conscience.

Addendum : États Mentaux sophistiqués

Conformément à l'objectif non révisionniste de montrer comment on peut retrouver et aborder les États Mentaux dans le cadre proposé ici, on a considéré jusqu'ici la notion ordinaire de croyance plutôt que celle, sophistiquée, proposée par des théoriciens tels Ramsey ou Lewis (§1.3.2). Afin de souligner les aspects importants de la critique des approches traditionnelles (ordinaires ou sophistiquées) du mental présentée dans §1.4, laquelle met en question l'efficacité de ces approches *pour aborder des questions détaillées et épineuses* telles celle de la fiction, il convient de mettre en évidence quelques points d'opposition entre l'approche proposée ici et les conceptions sophistiquées des États Mentaux. Outre la question des données qui soutiennent l'attribution des États Mentaux, ils sont en nombre de deux.

Premièrement, un aspect de la critique de §1.4 met l'accent sur l'imprécision des États Mentaux, laquelle constitue un embarras lorsqu'il s'agit d'aborder les questions détaillées et complexes relatives à la fiction et son rapport à la réalité (dans la psychologie du lecteur). On voit maintenant que le genre de résolution préconisée ici consiste à trouver un niveau plus détaillé, non pas sur le plan du comportement appréhendé par un observateur, mais plutôt sur celui de la conscience du sujet, de ses comportements en tant qu'il les appréhende, pour ainsi dire. *Cette critique s'applique aussi bien aux États Mentaux ordinaires qu'aux États Mentaux sophistiqués.* Elle met notamment l'accent sur les différences « d'activité » entre États Mentaux, que l'on reconnaît maintenant comme étant en rapport avec la *localité* de la structure instantanée de la conscience, et le fait que les éventuels aspects correspondant à certains États Mentaux y sont *hors sujet*. Alors qu'elle n'affirme que très peu d'explicite, la notion *ordinaire* d'État Mental semble pouvoir rendre compte *en théorie* du *hors*

sujet, dans la mesure où les blocs d'attribution relatifs à ces États Mentaux sont non saturés (chaque énoncé du langage n'est pas nécessairement accordé un statut de croyance) et la base d'attribution est locale (elle comporte certains comportements du sujet, mais non pas tous)⁹¹. En revanche, la notion *sophistiquée* d'État Mental écrase la notion de *hors sujet*, et cela pour deux raisons. D'un côté, le fait pour les blocs d'attribution des États Mentaux sophistiqués sont saturés (tout énoncé du langage est accordé un statut relatif au « mental », tel un degré de croyance) conjoint au fait que l'ensemble des États Mentaux dans le bloc entrent *d'emblée* en rapport avec le comportement du sujet (le « calcul » de l'action de l'agent dépend de tous les degrés de croyance de tous les énoncés)⁹², ont pour conséquence que chaque croyance de l'agent est *de prime abord toujours en jeu*. De l'autre côté, puisque les États Mentaux sophistiqués ont une base d'attribution globale, comportant notamment les comportements « *contrefactuels* » du sujet (§1.3.2), on compte au nombre des comportements auxquels il s'agit de rendre compte les réponses qui *auraient été* faites par le sujet si certaines questions lui avaient été posées. Ainsi, l'attribution des États Mentaux d'un Ramsey ou d'un Lewis prétend prendre en charge, par exemple, les réponses hypothétiques que le sujet aurait faites, alors même qu'il est en train de fuir éperdument devant un tigre, à la question de savoir combien fait sept plus cinq. Cette particularité, à part le fait de parvenir à une attribution saturée par une autre voie, a pour conséquence que les aspects que l'on qualifierait de *hors sujet* (les faits d'arithmétique lors de la fuite) sont mis explicitement en jeu, quoique seulement dans une situation contrefactuelle. En sorte d'écraser la distinction entre les États Mentaux *hors sujet* à un moment et ceux qui sont en jeu.

Un deuxième point d'opposition entre l'approche préconisée ici et le « raffinement » des États Mentaux ordinaires proposé par la théorie de la décision a trait à la question importante et complexe de la dynamique, laquelle a été reconnue comme la facette la plus épineuse de la question de la fiction et de son rapport à la réalité (§§1.4 et 1.5). Dans l'approche développée ici, la tâche de comprendre la dynamique est valorisée sur la tâche de rendre précise la notion de croyance ; en revanche, la théorie de la décision commence par rendre précise la

⁹¹Voir §1.3.2 pour une discussion approfondie de ces concepts.

⁹²Ce point général, fait dans §1.3.2, est d'autant plus important dans le cas du bloc posé par la théorie de la décision, à cause de son « compacité ». Dans la théorie de la décision, on attribue des degrés de croyances aux « mondes possibles » (états d'affaires maximaux) ; le degré de croyance à un énoncé est la somme des degrés de croyance aux mondes où il est vrai. L'action maximise l'utilité espérée, qui est calculée sur la base des degrés de croyance des mondes. Voir par exemple Joyce (1999), ou également §1.4.2, surtout sa note 94.

notion de croyance, notamment en termes de probabilités, pour ensuite essayer de penser leur dynamique à partir de ces notions précises⁹³. Cette différence n'est pas sans rapport aux accents différents mis sur la notion de *hors sujet* : on souligne ici l'importance des aspects immédiats de l'expérience, dont le *hors sujet*, *précisément en raison de leur importance pour la dynamique*, alors que le théoricien de la décision ignore ou minore ces aspects, et notamment le *hors sujet*, *précisément en raison du privilège accordé à la tâche de rendre précise la notion de croyance*. Naturellement, ceci mène dans deux directions différentes. Dans le cadre proposé, la question de la dynamique est *d'abord* la question de la dynamique de la conscience, à partir duquel on pourrait aborder *ensuite* la dynamique des États Mentaux. Dans l'optique de la théorie de la décision, la question est celle de la dynamique des États Mentaux sophistiqués, question qui ne se pose qu'après que les croyances (et les désirs) sont rendues précises en les accordant des « degrés ».

La différence entre les deux approches est mise clairement en relief par leurs attitudes différentes à l'égard du sens, normalement imprécis, de la *disposition* que l'on associe avec la croyance. Dans le cadre développé ici, comme l'a été remarqué ci-dessus, le « sens » de cette notion fuyante ne pourrait être proprement traité, approfondi ou précisé que dans la dynamique des structures instantanées de la conscience (pour autant que ce soit possible). À l'opposé, la théorie de la décision donne un sens précis à la disposition, en lui associant un nombre réel – le degré de croyance. Ce « raffinement » a des conséquences pour le rapport à l'action : une fois que l'on pose un degré de croyance pour « rendre précise » une disposition, on ne saurait faire le rapport avec le comportement autrement qu'en précisant *tous* les degrés de croyances pour *tous* les énoncés du langage, et *toutes* les utilités (degrés de préférence), et en les mettant *tous* en jeu, selon le principe de maximisation d'utilité espérée⁹⁴. Pour reprendre l'exemple où le sujet prend son parapluie en sortant, et particulièrement le moment où les objets *il pleut* et *prendre le parapluie* figurent dans la structure instantanée de la conscience avec une relation entre eux, la théorie de la décision implique une interprétation de cette relation selon laquelle elle est associée *non seulement* à la croyance que le parapluie protège de la pluie, *mais aussi bien* au désir de ne pas se mouiller. *Et bien*, vu l'exemple ci-dessus où il ne prend pas le parapluie, à la croyance qu'il ne l'a pas perdu. *Et bien* à un certain rapport entre le degré

⁹³Voir Lewis (1973b, p88) pour une affirmation explicite de cette méthodologie, quoique dans le contexte d'un débat différent.

⁹⁴On choisit l'action qui maximise l'utilité à laquelle l'on pourrait espérée étant donné ses croyances. Voir par exemple Joyce (1999), la note 92 ci-dessus ou la note 94 de §1.4.2.

de croyance qu'il le perdra et le désir de ne pas le perdre ... Bref, le « raffinement » de la notion de disposition proposé par la théorie de la décision creuse un écart entre l'expérience phénoménologique (la structure de la conscience) et l'interprétation complexe de cette expérience (l'interprétation de la relation entre les objets *il pleut* et *prendre le parapluie*). Dans une seule relation de la structure sont « recélées » un grand nombre de pensées (que le parapluie protège de la pluie, que le sujet ne veut pas se mouiller, que le parapluie n'est pas perdu, ...), lesquelles correspondent à grand nombre d'États Mentaux, en l'occurrence des croyances et des désirs : voilà l'image proposée par la théorie de la décision. Aucun de ces États Mentaux n'est *hors sujet* ; la relation entre ces deux objets est interprétée comme la présence d'eux tous.

Non seulement il s'agit d'une interprétation quelque peu tordue, sinon carrément forcée, de la situation mais, encore pire, cette conception se trouve fameusement dans l'embarras. Pour s'en apercevoir, il suffit de considérer une situation où le sujet remarque la pluie, mais sort précipitamment en oubliant son parapluie. En termes des structures instantanées de la conscience, cela se comprend assez facilement, tantôt comme l'absence de toutes les structures d'un objet tel *prendre le parapluie* et donc de relation avec *il pleut*, tantôt comme sa présence dans une structure instantanée lorsqu'il voit la pluie, mais au bord de la structure, pour ainsi dire, suivie par l'évanouissement dans les structures suivantes. Dans un cas, il ne pense pas à son parapluie (ni à ses propriétés protectrices), dans l'autre cas, le parapluie entre brièvement à l'esprit, pour ainsi dire, pour être immédiatement chassé par d'autres pensées plus urgentes. Dans les deux cas, le cadre développé permet d'expliquer son comportement par le *hors sujet* de la croyance que le parapluie protège de la pluie – cette croyance ne figure pas dans la structure instantanée de la conscience au bon moment (et dans la bonne façon). En revanche, la théorie de la décision se trouve dans un double embarras. D'une part, elle ne voudrait pas renoncer au fait que le sujet ait cette croyance, ni au fait qu'elle soit pertinente dans cette situation, d'où il doit soutenir sa présence, contrairement à son apparent caractère de *hors sujet*. D'autre part, il lui faut néanmoins expliquer le comportement du sujet, ce qui ne pourrait se faire que par l'ajout d'une autre croyance ou un désir qui « annule » la conséquence ordinaire de la croyance « oubliée », à savoir l'action de prendre le parapluie. Non seulement s'agit-il d'un ajout supplémentaire et artificiel d'un autre État Mental, mais il est difficile de voir quel État Mental ferait l'affaire : comment re-exprimer le fait d'oublier quelque chose en termes des croyances et des désirs seulement ?⁹⁵

⁹⁵Ce problème avec les États Mentaux proposés par la théorie de la décision s'adjoint à

Bref, la stratégie de préciser d'abord les notions de disposition et de croyance en termes des probabilités, et donc d'écraser le *hors sujet*, semble non seulement manquer largement les complexités des rapports entre les États Mentaux à un moment donné, et *a fortiori* leurs changements, mais elle échoue également face à un certain nombre de situations quotidiennes.

Peut-être que l'opposition entre l'approche développée dans ce travail et celle qu'emprunte les théories des États Mentaux sophistiqués, telle la théorie de la décision soit mise en relief de manière plus frappante dans un commentaire de l'un des moments fondateurs de celle-ci, à savoir le moment de *Vérité et Probabilité* où Ramsey pose et précise son problème de « mesurer les croyances »⁹⁶. Il commence (p162) :

Considérons donc maintenant ce qui est impliqué dans la mesure des croyances. Un système satisfaisant doit en premier lieu assigner à n'importe quelle croyance un grandeur ou un degré occupant une position définie dans un ordre de grandeur ; les croyances qui sont de même degré qu'une autre croyance, doivent elles-mêmes être de même degré et ainsi de suite.

Ce que Ramsey décrit *n'est pas*, d'après l'approche recommandée ici, la question de la mesure des croyances, car Ramsey fait une supposition douteuse. À savoir, il pense l'échelle qu'il faut utiliser pour mesurer des croyances à la façon des *nombre réels* : mesurer une croyance consiste à lui accorder un nombre réel (entre 0 et 1). L'analogie qu'il a en tête est celle avec la physique : de même qu'on mesure la vitesse en lui attribuant un nombre réel, on mesurerait la croyance en lui attribuant un nombre réel. Pour déjouer cette analogie, on peut en mobiliser une autre, tirée également de la physique. Le problème de mesurer la

d'autres problèmes bien connus ; de manière générale, les notions introduites ici, dont notamment la structure de la conscience et le *hors sujet*, ouvrent une voie prometteuse vers une résolution de ces problèmes. Par exemple, l'omniscience logique – le fait pour un sujet de croire toutes les conséquences logiques de ses croyances (Hintikka, 1975) – n'est nullement impliquée dans un cadre où un grand nombre de conséquences logiques d'une croyance sont *hors sujet* la plupart du temps (ou même toujours). D'un autre côté, les phénomènes rassemblés sous le nom du *framing problem* – contrairement à la rationalité parfaite, des actions des sujets dépendent apparemment de la *présentation* de la situation auxquelles ils font face, aussi bien que ses propriétés « logiques » (Tversky and Kahneman, 1981, 1986) – sont apparemment compréhensibles dans ce cadre en termes des aspects des structures instantanées de la conscience, dont leur forme (concentration sur un aspect de la situation au profit de l'autre) et leur localité (certains aspects pertinents selon la rationalité parfaite sont *hors sujet*).

⁹⁶Ramsey (2003, pp162-164).

croissance, suggère-t-elle l'approche développée ici, n'est pas analogue à celle de mesurer la vitesse mais plutôt à celle de mesurer le *mouvement*. Or, objecte-t-on, on peut mesurer la vitesse, l'accélération, et ainsi de suite, mais il ne fait pas sens de parler de « mesurer le mouvement » ! Exactement : en réponse à la question du mesurage du mouvement, on factorise (et précise) le mouvement en quelques éléments (vitesse, accélération, *etc.*) auxquels on peut associer des nombres réels ; c'est-à-dire des éléments que l'on peut mesurer dans la manière que Ramsey avait en tête. Dans ce sens le mouvement ne se mesure pas par *un* nombre réel, mais par une suite de nombres réels : l'échelle de mesure n'est pas *uni-dimensionnelle* (comme pour la vitesse), mais *multi-dimensionnelle*. Dans le sens où l'on préconise ici la considération – ou si l'on préfère le « mesurage » – des croyances relatives à certaines successions de structures (les structures instantanées de la conscience), on recommande une échelle de mesure plus complexe que celle que Ramsey supposait⁹⁷.

Deux alinéas plus tard (p163), Ramsey discute une option pour effectuer cette mesure :

Nous pouvons, en premier lieu, supposer que le degré d'une croyance est quelque chose de perceptible par son possesseur, c'est-à-dire que les croyances diffèrent par l'intensité d'un sentiment qui les accompagne, que nous pourrions appeler un sentiment-de-croyance (*belief-feeling*) ou sentiment de conviction, et par degré de croyance nous voulions dire l'intensité de ce sentiment (*feeling*). Cette manière de voir ne conviendrait cependant pas du tout parce qu'il n'est pas facile d'assigner des nombres à des intensités de sentiment. Même cela mis à part, ça me semble visiblement faux puisque les croyances auxquelles on tient le plus fortement ne sont souvent accompagnées d'aucun sentiment que ce soit ; personne ne tient à cœur ce qui est pris pour acquis.

Par suite, il écarte cette option au profit de celle que l'on a déjà discutée et qui lui est associée⁹⁸. Il est désormais clair que l'approche développée ici n'entend nullement identifier ni « mesurer » les croyances directement par des expériences phénoménales immédiates du sujet ; il est également clair que l'approche ne conçoit aucun sentiment spécifique qui accompagne les croyances, aucun « sentiment-de-croyance ». En revanche, l'approche soutient que l'expérience phénoménale entre dans un rapport important avec la croyance, laquelle est as-

⁹⁷ Au lieu d'employer la structure « linéaire » des nombres réels, on emploie effectivement la structure plus libre des successions des structures instantanées de la conscience.

⁹⁸ « Le degré de croyance est une propriété causale de la croyance, qu'on peut vaguement définir comme étant la propension à agir sur la base de cette croyance » (p164).

sociée aux propriétés des successions des structures instantanées de la conscience ; également, elle soutient que la croyance pourrait avoir une « présence » dans la conscience *lorsqu'elle est en jeu* sans correspondre à aucun sentiment explicite et remarqué. Que l'on s'attarde sur la manière dont cette approche répond aux deux arguments de Ramsey.

Premièrement, Ramsey soutient que le problème avec les sentiments est la difficulté de les mesurer. On a déjà répondu à cet argument : on peut « mesurer » l'expérience phénoménologique à un instant, ou plus précisément l'état instantané de la conscience, mais avec une échelle de mesure plus subtile et sophistiquée que les nombres réels, à savoir les structures au sens de §2.1.3.

Par ailleurs, le deuxième point de Ramsey est ambigu entre deux possibilités. D'une part, il pourrait vouloir mettre en évidence le fait que parfois des croyances sont « en jeu » sans être focalisées ou remarquées (dans la mesure où ce qu'il appelle sentiments sont munies de degrés d'intensité, il semble que des croyances « accompagnées d'un sentiment » sont des croyances plus ou moins « explicitement remarquées »). L'approche développée est d'accord sur ce point, mais souligne le fait qu'une croyance pourrait être « présente » dans un état instantané de la conscience dans d'autres manières : il pourrait, pour parler comme Ramsey, entraîner des « sentiments » sans qu'ils soient reconnus explicitement comme associés à la croyance. Dans les termes introduits ci-dessus (§2.2.2), la croyance peut être non focalisée. Selon l'autre interprétation du deuxième point de Ramsey, il voudrait mettre le doigt sur les croyances qui *ne sont pas* « en jeu » à un moment, qui donc ne correspondent à aucun « sentiment » à ce moment, mais qui néanmoins *devraient être accordées un degré*. C'est tout évidemment l'attestation même d'une dévalorisation du *hors sujet*, laquelle on a suffisamment considéré déjà.

Dans un certain sens (et, bien entendu, hors la question des données que l'on se permet pour « mesurer » des croyances), l'approche préconisée ici au mental, et la conception qu'elle implique des croyances est la synthèse des deux options considérées par Ramsey. D'un côté, en associant des croyances à des propriétés des successions des structures instantanées de la conscience, elle rejoint le dynamisme de la notion de croyance comme propension à agir, et donc « mesurable » ou « apercevable » seulement dans ces actions, ou plus généralement dans des suites d'événements (comportant ce qui arrivent au sujet et ses actions). De l'autre côté, en pensant la croyance à partir des états instantanés de la conscience, exprimés de manière rigoureuse, et notamment en évoquant le rapport entre la dynamique continue des états instantanés et leurs aspects intérieurs, on rend justice à l'intuition d'un rapport entre l'expérience

phénoménale et les croyances, quoiqu'un rapport qui ne soit ni constitutif ni certain.

La croyance est comme la vitesse moyenne d'un mobile. Dans les premières sections de ce chapitre, on a caractérisé l'état instantané du mobile – la structure instantanée de la conscience. En prenant la succession des états instantanées on pourrait retrouver la vitesse moyenne comme une propriété s'appliquant à cette succession – la croyance comme propriété de la succession de structures instantanées. En revanche, la théorie de la décision, dans l'engrenage ramseyen, essaie de caractériser et de travailler avec la vitesse moyenne – la croyance – directement. L'une de ces deux approches est évidemment plus féconde pour penser la dynamique du mobile en général.

Chapitre 3

Retrouver la fiction

Ce chapitre, qui conclut la première partie de cette thèse, a pour but d'entamer une théorie de la fiction et de son rapport à la réalité, en tant qu'elle apparaît dans la psychologie du lecteur. Pour ce faire, il s'agit de mobiliser d'une part les résultats de la théorie développée précédemment relative aux états instantanés de la conscience, à son dynamique, et aux États Mental, et d'autre part, l'hypothèse admise dans §1.1.2 que le lecteur sait que le texte qu'il lit est fictionnel. La stratégie consistera, non pas à poser un processus différent pertinent pour la lecture ou le traitement de la fiction, ni à poser un royaume mental consistant en des entités à titre particulier (les entités « fictionnelles »), mais simplement à traiter la connaissance que le texte est fictionnel comme toute autre connaissance ou croyance¹ et à lui appliquer la théorie développée dans le chapitre précédent. En tant que croyance, elle est d'abord associée à des propriétés des successions des structures instantanées de la conscience où elle est « en jeu » ou « effective ». Par suite, et en conséquence de la continuité de la dynamique des structures de la conscience, elle pourrait « correspondre » à certains objets, relations ou aspects de certaines structures instantanées de la conscience à certains instants (§2.5.3). Les aspects, propriétés, capacités et complexités de la fiction (en tant qu'elle apparaît dans la psychologie du lecteur) se comprennent, pour autant qu'ils soient effectifs, sur la base de ces successions, de ces structures instantanées de la conscience, et de leurs propriétés. En telle sorte qu'une réponse à la question de la fiction consisterait dans une *explicitation* des successions appropriées à la question de la fiction et une *constatation* des propriétés pertinentes et intéressantes de ces successions. Comme on a souligné lors de l'explicitation de cette stratégie « naturaliste » dans §1.2.3, par

¹L'acception du terme « croyance » ici, héritée de la philosophie de l'esprit, tend à recouvrir la connaissance aussi (voir par exemple Dennett (1987, p46, note 2)).

cette démarche on *retrouve* la distinction stricte entre fiction et réalité que soutiennent ou supposent d'autres théories de telle manière que l'on comprend en même temps et de façon naturelle les difficultés auxquelles elles se heurtent.

En général, les aspects intéressants relatifs à la fiction, et les questions à son égard, se divisent en deux, à savoir, ceux qui peuvent être considérés dans les états instantanés de la conscience individuels, à des moments particuliers, et ceux qui appellent un recours aux propriétés des successions des états instantanés. Vu le progrès que l'on a fait sur la question de l'état instantané de la conscience (§§2.1 et 2.2), on saurait caractériser et comprendre un grand nombre des aspects qui se pensent dans des états instantanés individuels. À cette fin, la *localité* de l'état instantané de la conscience, ainsi que sa structure de *focalisation*, seront des outils indispensables, permettant non seulement de caractériser des propriétés de la fiction, mais également de tirer quelques conséquences intéressantes à son égard. En revanche, quoique le côté dynamique de la théorie du mental soit bien fondé, il est peu développé, si bien que des conclusions puissantes ne peuvent en être tirées à l'heure actuelle pour ce qui concerne les aspects de la fiction. On se contente dans ce qui suit d'identifier les aspects, propriétés et problèmes de la fiction qui renvoient à la dynamique, sans s'y attarder plus. Ainsi, on présente autant une vision de la fiction en tant qu'elle figure dans la psychologie du lecteur, en l'occurrence la vision impliquée par la théorie du mental développée précédemment, qu'une théorie exhaustive.

En conséquence de la *continuité* de la dynamique de la conscience, c'est-à-dire du *rapport étroit et important* entre l'état instantané de la conscience et sa *dynamique* (§2.3), les questions relatives aux états instantanés de la conscience à des instants particuliers sont parfois fortement reliées aux questions relatives à leur dynamique. C'est effectivement le cas pour deux des « facettes » que l'on a reconnues aux questions **A.** et **B.** de §1.1.1, à savoir la question *des conduites du lecteur dans lesquelles il montre une « gestion » du rapport entre réalité et fiction*, de ce que l'on a appelé dans §1.1.3, à l'instar de Schaeffer, la « *compétence fictionnelle* », et la question de *son état cognitif et affectif à des instants particuliers pendant ces « conduites »*, instants où, on l'a vu dans §1.4.1, il y a des rapports apparemment complexes entre le « côté fictionnel » et le « côté réel ». La première question relève de la dynamique, alors que la deuxième a trait aux états instantanés particuliers. Néanmoins, la continuité de la dynamique implique un rapport étroit entre elles, lequel pourrait être mobilisé dans l'un sens et dans l'autre. D'une part, des aspects dynamiques suggèrent que les états instantanés ont certaines propriétés plus ou moins intéressantes, et, d'autre part, la question dynamique pourrait être traitée dans une large mesure en s'ap-

puvant sur l'étude des aspects des structures instantanées de la conscience à certains instants particuliers. Dans la première section de ce chapitre, on considère quelques compétences fictionnelles (l'immersion fictionnelle, la sortie et l'entrée de l'immersion fictionnelle, l'évitement de confusions entre fiction et réalité) et les états instantanés de la conscience à des instants appropriés (instant d'immersion fictionnelle, d'entrée ou de sortie et ainsi de suite), lesquels ont été identifiés comme difficiles dans §1.4.1. L'étude de ces états instantanés ne débouche pas seulement sur des résultats et des conséquences intéressants en soi – plus notamment la caractérisation d'un état d'immersion fictionnelle comme un état où le fait qu'il s'agit d'une fiction est *hors sujet* –, mais devrait être également considérée comme une contribution à la compréhension générale des conduites du lecteur dont est composée sa compétence fictionnelle.

En revanche, lorsque l'on vient, dans la deuxième section de ce chapitre, à considérer les questions **C.** et **D.** de §1.1.1 ayant trait aux éventuels transferts ou chevauchements entre la fiction et la réalité, la stratégie d'aborder la dynamique à partir d'une étude des états instantanés de la conscience s'avérera moins féconde. Ces questions ont été reconnues pour plus complexes et difficiles, dans la mesure où elles ont affaire non seulement aux aspects du mental du lecteur ayant trait à la fiction, mais aussi à d'autres aspects « ordinaires » ou « réels » de son mental, dont ses préjugés, ses croyances, ses avis : c'est-à-dire, non seulement aux propriétés des successions des états instantanés de la conscience ayant trait à la fiction, mais également aux propriétés des successions « usuelles », et aux rapports entre elles. Par conséquent, la réponse pleine à la question des croyances réelles mises en jeu dans la lecture de la fiction, ainsi qu'à celle des aspects fictionnels qui s'infiltreraient parmi les croyances à l'égard de la réalité, requiert une compréhension précise de *la totalité de la dynamique* des états instantanés de la conscience. Bien que l'on s'attarde brièvement sur quelques états instantanés intéressants, il s'agira pour l'essentiel de remarquer que la question renvoie à la dynamique de la conscience.

Pour finir, on remarque que ni les États Mentaux spécifiques à la fiction (*faire comme si*, imagination), ni les problèmes posés par les rapports avec des États Mentaux usuels (croyances, émotions), n'échappent à l'approche proposée. En effet, les États Mentaux proposés comme spécifiques à la fiction sont des États Mentaux *narratifs* (dont la base d'attribution couvre plusieurs moments, §1.3.2), et donc associés à des successions des états instantanés de la conscience plutôt qu'à des états instantanés individuels à des instants particuliers. Comme pour le cas des croyances (§2.5.3), où l'on comprend les complexités, les aspects et la dynamique de ces États Mentaux et, notamment, des rapports entre eux

à partir de la dynamique des états instantanés de la conscience eux-mêmes, on arrive à comprendre les supposés États Mentaux spécifiques à la fiction et leurs rapports problématiques aux croyances en termes des états instantanés de la conscience. On considérera brièvement ces questions dans la troisième et dernière section du présent chapitre.

3.1 Compétence fictionnelle

Le thème de la lecture comme démarche, comme processus dynamique, est probablement aussi vieux que la forme artistique, du moins dans son versant littéraire ou poétique, elle-même². Dans ce travail, la démarche appropriée est celle des structures instantanées de la conscience du lecteur. Or, plutôt que de se limiter à sa lecture, il s'agit également de comprendre ses dits et ses comportements hors la lecture (dans la discussion du texte de fiction, par exemple), et cela encore en s'appuyant sur les successions des structures instantanées appropriées. Non seulement l'appréciation de la fiction, mais aussi sa gestion, sa discussion, l'évitement de la confusion avec la réalité pendant et après la lecture, sont tous à comprendre comme des aspects, des tendances, des régularités des successions des structures instantanées de la conscience. Dans ce cadre, la distinction entre fiction et réalité est à comprendre *d'abord* dans et par ces manifestations, ou si l'on préfère ces conséquences, dynamiques.

Il ne s'agit pas ici de développer une théorie ou une caractérisation complète des aspects dynamiques correspondant aux diverses compétences fictionnelles, mais plutôt d'aborder les structures instantanées de la conscience à des instants appropriés de ces passages dynamiques. D'une part, l'approfondissement de ces structures instantanées fournit des indications, pas infaillibles mais néanmoins pas inutiles, relatives à la succession des structures instantanées qui précèdent et qui suivent immédiatement. D'autre part, la théorie de la structure instantanée de la conscience développée précédemment permet de bien comprendre et caractériser certaines propriétés du rapport entre fiction et réalité à ces instants, qui ont par ailleurs posé problème pour d'autres théories de la fiction (§§1.4 et 1.5).

Dans cette section, on considère en gros trois sortes de compétence fictionnelle. Tout d'abord, la compétence fictionnelle de base, l'état « standard » de

²Le thème de la narrativité, de l'importance de la démarche de l'œuvre, est naturellement très présent chez Aristote (1990). Plus récemment, certains philosophes, tels Iser (1976) et Feagin (1996), ont mis l'accent sur le caractère dynamique de la lecture.

la lecture si l'on peut dire, est sans doute « l'immersion fictionnelle »³, c'est-à-dire l'état du lecteur lorsqu'il est pleinement « pris » par sa lecture, « investi » dans son livre, « perdu » dans la fiction. Toutes ces métaphores pointent une sorte d'*absence* des faits bruts de la réalité, dont le fait même qu'il s'agisse d'une fiction et, comme on l'a vu dans §1.4.1, les faits réels qui contredisent les propos du texte ; en effet, dans §3.1.1, on parvient à caractériser cet état par le fait que le caractère fictionnel du texte est *hors sujet* relativement à la structure instantanée de la conscience à ce moment. Il s'agira également de tirer quelques conséquences de ce *hors sujet* et de confronter d'autres thèses relatives au caractère du fait de fiction à des moments de la lecture.

Or, l'état d'immersion fictionnelle n'est nullement le seul composant de la compétence fictionnelle. Il faut également qu'il soit capable d'adopter un état d'immersion fictionnelle, ainsi que celle de s'en dégager ; d'ailleurs, on ne comprend pas l'immersion fictionnelle si l'on ne comprend pas les démarches qui y ont mené qui la suivent. Par ailleurs, on compte parmi ses compétences fictionnelles le fait de pouvoir prendre une vue de l'extérieur par rapport au texte fictionnel aussi bien que la vue d'intérieur, et le fait d'éviter des éventuelles confusions entre la fiction et la réalité qui se présentent aussi bien dans sa lecture que dans sa vie quotidienne. Si l'immersion fictionnelle, et bien sûr les moments « ordinaires » de la vie quotidienne, se caractérisent par le *hors sujet* du fait de fiction, alors, pendant les démarches relevant de ces autres compétences fictionnelles, où se mêlent la fiction et la réalité, le fait de fiction est « *en jeu* » d'une manière ou d'une autre. Par exemple, pendant les entrées « typiques » dans l'immersion fictionnelle – où l'on ouvre le livre, où le rideau s'élève dans le théâtre, où les lumières s'éteignent dans le cinéma – le fait de fiction est d'abord en jeu pour s'estomper dans l'ombre au fur et à mesure que l'on entre dans l'immersion. De l'autre côté, lors des sorties de l'immersion fictionnelles – lesquelles représentent des aspects importants de maîtrise de la distinction entre fiction et réalité de la part du lecteur – le fait qu'il s'agit de la fiction semble rentrer encore en jeu.

Une manière de trier ces compétences fictionnelles s'appuie sur le caractère *focalisé* ou non de la mise en jeu du fait de fiction (§2.2). Pour donner quelques exemples des cas où le fait de fiction est focalisé, on peut prendre premièrement les moments où le lecteur considère des énoncés portant *sur* le texte mais qui ne sont pas vrais *selon* le texte (tel « Holmes est un personnage fictionnel »). Dans ces moments, il aborde le texte *à la fois* du dedans et du dehors ; les structures instantanées de la conscience à ces instants comportent aussi bien des aspects

³Schaeffer (1999, Ch. III, §4).

que l'on qualifie de « fictionnels » que des aspects qu'on qualifie de « réels » et une distinction *focalisée* entre eux. Un autre exemple est celui de l'interruption de la lecture pour poser la question de l'existence de Holmes et de l'hésitation qui en résulte entre la réponse « fictionnelle » (il existe) et la réponse « réelle » (il n'existe pas), exemple qui a été reconnu dans §1.4.1 comme difficile pour les théories traditionnelles de la fiction, alors même qu'elle constitue un aspect important de la question **A.** de §1.1.1 et plus généralement du comportement du lecteur à l'égard de la fiction. Encore une fois, la structure instantanée de la conscience comportera, de manière complexe mais généralement *focalisée*, la distinction entre aspects « fictionnels » et « réels ». Un dernier exemple, ayant trait aux affects relatifs à la fiction et donc à la question **B.** de §1.1.1, est celui, relevé dans §1.4.1, où le lecteur tente de quitter – ou d'échapper à – l'immersion fictionnelle lors d'une épisode de sentiment éperdu, en se rappelant qu'il ne s'agit que de fiction. À un tel instant, le fait qu'il s'agit de la fiction est présent explicitement dans la structure instantanée, et il y entretient un rapport complexe avec le sentiment éprouvé.

Bien évidemment, tous ces cas ont trait à des moments dans des passages dynamiques qui constituent d'importants aspects du comportement du lecteur à l'égard de la fiction : la possibilité d'adopter un regard critique et d'en sortir, la capacité à répondre à une question dont la position interrompt la lecture, l'enjeu entre ses sentiments à l'égard de la fiction et ses connaissances à son égard. Dans tous les cas, pour comprendre entièrement le comportement – les modalités d'entrée et de sortie d'une perspective « critique », la réponse et l'hésitation devant la question, les modalités de mise en jeu du fait de fiction comme « défense » contre un sentiment, et l'efficacité de cette mise en jeu pour mitiger le sentiment accablant – il faut se tourner vers la dynamique des structures instantanées. Néanmoins, l'examen des structures instantanées particulières et du caractère de la présence du fait de fiction dans ces structures peuvent apporter quelques éléments à la compréhension de ces démarches et de ces capacités ; dans §3.1.2, on considère les trois exemples ci-dessus et plus généralement les mises en jeu *explicites* ou *focalisées* du fait de fiction, en s'appuyant sur les structures instantanées à des moments particuliers.

D'un autre côté, la « compétence fictionnelle » qu'a le lecteur d'éviter la confusion entre fiction et réalité, qu'elle soit pendant la lecture ou dans la vie « ordinaire » fournit des exemples des entrées en jeu *non focalisées* du fait de fiction. Par exemple, l'amateur de Sherlock Holmes qui a, au cours de sa vie quotidienne, sous ses yeux les titres de propriété de 221b Baker Street datant de la fin du dix-neuvième, ne voit aucune contradiction dans l'absence du nom

de Holmes. Cette situation met à l'épreuve la conception du lecteur à l'égard de la distinction entre fiction et réalité, et sa gestion de cette distinction, dans la mesure où elle offre la possibilité de confusion. Pour comprendre pleinement cette gestion, il faut naturellement aborder la dynamique de la succession des structures instantanées de la conscience à ces moments ; cependant, conforme à la stratégie générale adoptée dans cette section, on arrivera à comprendre certains aspects de cette succession en considérant en détail une structure instantanée à un moment important de la conduite du sujet. Comme on verra dans §3.1.3 lorsque l'on considère ce genre d'exemple en plus de détail, le fait de fiction est « présent » ou « effectif » dans cette structure, mais seulement dans un sens « structurel », non explicite, *non focalisé*.

3.1.1 L'état d'immersion fictionnelle

Où l'on analyse l'immersion fictionnelle, l'état où le lecteur « se laisse prendre » par le texte ou « s'oublie » dans sa lecture, comme l'absence du fait qu'il s'agit de la fiction par rapport à la structure instantanée au moment approprié. Cette absence instantanée permet d'expliquer la dite « apparence de réalité » de la fiction, aussi bien que le sens dans lequel le lecteur entre « dans » sa lecture, alors même qu'il garde la connaissance qu'il s'agit de la fiction. Pour terminer, on défend cette analyse contre certains auteurs qui ont soutenu la thèse opposée, d'une expérience phénoménologique spécifique à la fiction.

Dans §1.4.1, on a posé la question suivante : quel est le statut de la croyance que Holmes n'existe pas, à un moment où le lecteur est immergé dans sa lecture des *Aventures de Sherlock Holmes* ? Dans §1.4.2, on a posé une question semblable : quel est le statut de votre croyance (oubliée) relative à un camarade de classe, à un moment avant que la question vous soit posée ? La réponse simple à cette dernière question, donnée dans §2.5.3, est que cette croyance est absente de votre état instantané de la conscience à cet instant. Elle est *hors sujet*. De manière pareille, il semble qu'un grand nombre de croyances (et connaissances) ordinaires qui sont contredites dans la fiction soient simplement *absentes* de l'état instantané de la conscience du lecteur à des moments de (et par ailleurs après) sa lecture. L'opposition à la réalité sur ces points est tout simplement *hors sujet*.

En outre, le *hors sujet* constitue un apport clef pour la compréhension de l'état d'immersion fictionnelle. Intuitivement, on dit que, à un moment de l'immersion fictionnelle, le fait même qu'il s'agit d'une fiction est « laissé à côté »

ou même « oublié » dans un certain sens, et le lecteur se laisse pris par le jeu du texte. En appliquant la théorie rigoureuse des structures instantanées de la conscience, on obtient le résultat suivant : *le fait qu'il s'agit d'une fiction est tout simplement hors sujet, au sens où il est absent de la structure instantanée* (§2.2.1). Cette caractérisation de l'état d'immersion fictionnelle reçoit un soutien supplémentaire des considérations de §2.5.3 à l'égard des modalités de l'apparition des croyances dans la structure instantanée de la conscience à un instant. Comme on l'a souligné dans §2.5.3, le fait pour une croyance d'être « en jeu » ou « effective » dans un certain intervalle temporel a pour envers certains aspects dans certaines structures instantanées à des instants de cet intervalle, lesquelles « correspondent » à la croyance, si l'on peut dire. Par conséquent, dans la mesure où le fait qu'il s'agit de fiction – fait qui est l'objet d'une connaissance de la part du lecteur – est apparemment inopérant pendant de larges intervalles de la lecture, et surtout de l'immersion fictionnelle la plus « profonde », il n'y a pas d'aspects des structures instantanées à ces instants qui peuvent « correspondre » à la connaissance que le texte est fictionnel ; autrement dit, ce fait est *hors sujet* à ces instants.

Bien entendu, l'identification de l'immersion fictionnelle avec le *hors sujet* connaît des réserves habituelles relatives au caractère vague des structures instantanées de la conscience. En raison de ce vague, il n'est pas possible de dire que le fait de fiction est totalement absent, mais on pourrait au moins dire qu'il est tellement enfoncé dans l'arrière-plan à ne pas exercer d'effet important sur la structure de cette conscience. C'est-à-dire, si l'on pense qu'il y a un objet *fiction*, alors son appartenance à la structure est fortement indéterminée, et autrement, s'agissant d'une relation qui « correspond à » ou « recèle » prétendument le fait de fiction, sa présence dans la structure est également une question douteuse.

La reconnaissance du *hors sujet* du fait de fiction à des moments d'immersion fictionnelle s'avère fort utile, permettant de rendre compte de façon naturelle d'un certain nombre de phénomènes relatifs à la fiction. On s'attarde sur deux d'entre eux, à savoir la dite « apparence de réalité » de la fiction d'une part et les intuitions sous-jacentes à la théorie de fiction comme jeu de *faire comme si* d'autre part. Pourtant, elle s'oppose à la thèse d'un caractère spécifique à l'apparition d'un événement ou un aspect fictionnel dans la conscience, laquelle a été soutenue apparemment par Sartre, Ingarden, et, avec des arguments différents, Schaeffer ; il faut donc revenir, dans un deuxième temps, sur les arguments soutenant cette thèse.

Tout d'abord, dès que l'on reconnaît l'immersion fictionnelle comme l'ab-

sence du fait de fiction, il découle qu'une structure instantanée de la conscience à un instant d'immersion n'a pas de différence structurelle avec une structure instantanée à un instant « ordinaire non fictionnelle ». Ce qui explique la dite « apparence de la réalité » que l'on constate souvent relativement à la fiction⁴. En effet, non seulement le fait de ressembler à une structure instantanée « ordinaire » implique que la structure instantanée à un moment d'immersion ne comporte aucun aspect permettant de distinguer entre les événements ou personnages fictionnels et ordinaires, mais, de plus, puisque le sentiment éprouvé à un instant correspond à une propriété globale de la structure instantanée de la conscience à cet instant (§2.5.2), l'absence de différence structurelle entre une structure instantanée à un moment « ordinaire » et une structure à un moment d'immersion fictionnelle a pour conséquence qu'il n'y a pas de « sentiment de fiction » nécessairement différent du « sentiment de réalité »⁵. Suite à l'analyse de l'immersion fictionnelle comme *hors sujet* du fait de la fiction, l'apparence de réalité, qu'il s'agisse des événements décrits ou de l'expérience de lecture, s'avère n'être rien d'autre que « l'investissement » plein dans la lecture et dans la situation que l'on est en train de lire, « l'investissement » à tel point de perdre vue du fait que c'est de la fiction. Par conséquent, l'immersion fictionnelle n'est pas d'un genre différent de tout autre « investissement » où l'on s'occupe pleinement d'une seule situation, telle lorsque l'on concentre sur une chose au détriment de ce qui l'entoure ou lorsque l'on porte son attention sur les événements qui se déroulent dans une salle et ignore tout ce qui se passe hors la fenêtre. Conséquence également immédiate, l'apparence de réalité n'a rien à voir avec le genre du texte, car le fait que le texte est fictionnel pourrait bien être absent dans la lecture de n'importe quel texte de n'importe quel genre (le fait pour certains genres littéraires d'être plus ou moins habiles à guider la conscience du lecteur dans un tel état est une autre question, relevant au moins partiellement de la dynamique).

Par ailleurs, la thèse du *hors sujet* du fait de fiction aux moments d'immersion fictionnelle rend compte des deux grandes intuitions qui soutiennent et motivent le rapprochement entre la fiction et le jeu de *faire comme si* proposé par Walton et Currie (§1.2.1), sans rencontrer les difficultés qui ont touchées à leur théorie. D'un côté, il y a l'idée que *dans* le jeu de *faire comme si* (et également, *dans* la lecture de la fiction), le fait que ce soit du *faire comme si*

⁴Par exemple, Ingarden (1968, p52) : « le caractère ontique les objets représentés dans un œuvre d'art littéraire est généralement celui de la réalité ».

⁵Il ne s'agit évidemment pas à nier que les fictions puissent susciter des sentiments que l'on n'éprouvent pas ordinairement dans la réalité, mais de les comprendre comme le propre du bon écrivain, non pas celui de la fiction elle-même.

est absent (d'où la ressemblance entre l'intérieur du jeu et la réalité). De l'autre côté, un jeu de *faire comme si* (ou autrement dit, le fait qu'il s'agit de fiction) est *persistant* dans le sens où, même lorsque l'on est dedans, il est toujours « là ». Plus : il faut qu'il soit « là » pour que l'on puisse être dedans. Walton et Currie ont peine à expliciter le rapport entre l'intérieur du *faire comme si*, où le fait de fiction est *absent*, et le fait *persistant* qu'il s'agisse du *faire comme si*, c'est-à-dire sa *présence* ; c'est d'ailleurs l'une des sources des problèmes relevés dans §1.4.1. Par contraste l'approche proposée ici résout cette tension grâce à la *localité*. Le fait que c'est du *faire comme si* ou de la fiction est bien *absent* aux moments d'immersion, en l'occurrence des structures instantanées *locales* à ces moments (§2.2.1). S'il est « *présent* », c'est dans un autre sens : il s'agit d'une présence à la façon d'une croyance qui, alors qu'elle est attribuée au sujet sur des longues périodes, demeure longtemps « latente », pour n'entrer en jeu que dans des conduites précises et temporellement locales. Il y aura des aspects « correspondant » à cette croyance seulement dans les structures instantanées de la conscience à des moments où elle est en jeu, et certainement pas à chaque moment de la vie du sujet. De manière semblable, le fait qu'il s'agit de la fiction est bien « là » en tant que croyance du lecteur, mais elle entre en jeu et est « présent » en tant qu'aspect des structures instantanées seulement aux moments où le rapport entre fiction et réalité est en jeu et non pas « en pleine lecture ». C'est le dépassement de la simple dichotomie entre le *dans* et le *hors* du *faire comme si* au profit de la gamme de différents « *statuts* » possibles du fait de fiction, selon qu'elle est objet ou relation d'une structure instantanée, selon que sa présence y est plus ou moins vague, selon qu'elle s'associe à certaines structures instantanées plutôt que d'autres, qui permet au cadre proposé d'expliquer les intuitions des théories précédentes sans subir leurs inconvénients.

La thèse de l'*absence* (locale) du fait de fiction à des moments d'immersion fictionnelle est la conséquence d'une application simple de la théorie du mental développée dans le chapitre précédent aux observations quotidiennes relatives à l'expérience de lecture. Néanmoins, elle s'oppose à une thèse d'une phénoménologie spécifique à la fiction, que l'on peut repérer, dans diverses formes, aussi bien chez des phénoménologues tels Sartre et Ingarden que chez des théoriciens de la fiction tel Schaeffer. Quant aux premiers, ils soutiennent des aspects de la conscience qui sont spécifiques à la fiction : Sartre soutient une distinction entre la « conscience » spécifique à la fiction (ou l'imagination) et d'autres « consciences », et Ingarden, pour sa part, soutient que les énoncés d'un texte de fiction sont des « quasi-jugements », requérant des « actes de conscience » ou des « moments « thétiques » » particuliers de la part du lec-

teur⁶. Pour formuler leur thèse dans le langage du cadre mental proposé ici, elle affirme qu'il y a une relation unaire « ... est fictionnel » dans toute structure instantanée de la conscience, qui s'applique à tous les objets de la structure qui sont « fictionnels » et seulement à ces objets⁷. Il pourrait sembler que le débat sur la présence ou non d'une telle relation dans une structure instantanée consiste en un simple *petitio principii*, d'autant plus que le vague de la structure instantanée empêche de trancher une fois pour toute la question. Or, comme on l'a vu dans §2.3, les propriétés d'une structure instantanée particulière sont fortement reliées au passage dynamique dont cette structure fait partie, de telle sorte que l'importance, sinon la présence, d'une relation particulière dans une structure instantanée pourrait être jugée selon ses conséquences dynamiques. D'ailleurs, un des arguments en faveur du rapprochement entre l'immersion fictionnelle et l'*absence* du fait de fiction s'appuie précisément sur le manque de conséquence « effective » de la connaissance au fait de fiction sur la dynamique de la lecture, notamment à des moments d'immersion fictionnelle. Il s'ensuit que même si la relation « ... est fictionnel » soutenue par Sartre et Ingarden, et donc la distinction qu'elle fait entre les aspects « fictionnels » et « non fictionnels », est présente dans une structure instantanée de la conscience à un moment d'immersion fictionnelle, elle l'est de manière tellement à l'arrière-plan et tellement inefficace qu'il n'y a aucune différence avec le cas où elle est absente. La thèse de l'*absence* du fait de fiction à des moments d'immersion fictionnelle, nuancée, comme on l'a déjà remarqué, par le vague des structures instantanées, n'est donc pas mise en question.

D'un autre côté, Schaeffer paraît rejeter la thèse de l'*absence* du fait de fiction à des moments d'immersion fictionnelle, lorsqu'il prétend identifier un aspect *phénoménologique de base* de l'immersion fictionnelle. Il s'agit en l'occurrence de « l'attention scindée », qui « aboutit à l'existence de deux mondes,

⁶Sartre (1995) soutient que « les phrases du roman se sont imbibées de savoir imageant », où « le savoir imageant [à la différence du « savoir pur » et du jugement] est une conscience qui cherche à se transcender, à poser la relation comme un *dehors* » (p132, voir en général pp128-134). Ingarden (1968) prétend que les « moments « thétiques » spéciaux [impliquées dans l'appréhension des quasi-jugements d'un texte de fiction] apparaissent ici comme une variation particulière des moments de reconnaissance de l'existence qui apparaissent dans l'appréhension d'un objet réel » (p214, voir également p215, note 40, et §13). Ce qui importe pour la discussion présente est la thèse d'une « conscience » ou d'un « acte de conscience » particulier à la fiction, et non pas les thèses (d'ailleurs partagées par les deux auteurs) relatives à la particularité des énoncés méritant ce genre de « conscience » ou aux autres conséquences (relatives « au rôle des signes » ou à « l'expérience esthétique » par exemple) de ce genre de conscience.

⁷Cf. la note 73 de §2.5.1, où l'on a rapproché la thèse d'une relation unaire « ... est une pensée » dans toute structure instantanée de la thèse d'une conscience irréfléchie.

celui de l'environnement réel et celui de l'univers imaginé »⁸. Pourtant, « l'attention scindée » de Schaeffer semble renvoyer à une différence entre les *entités du texte* et les éléments de l'*environnement réel* immédiat, c'est-à-dire une différence entre le « dans le texte » et « dans l'environnement de lecture ». Cette différence n'a nullement particulier à la fiction ; au contraire, elle a seulement affaire au fait que ce qui est décrit dans le texte ne déroule pas simultanément sous les yeux du lecteur, fait approprié à la lecture de la plupart de textes (sinon tous), qu'ils soient fictionnels ou non. De même que Giono pourrait se figurer les personnages de sa fiction dans son environnement immédiat, sans « confondre » personnage fictionnel et environnement réel (« Dans ma fenêtre ouest, j'ai installé ... Mme Tim ... »⁹) ; de même Michelet pourrait se figurer les grands personnages de la Révolution dans son environnement immédiat (Robespierre dans sa fenêtre ouest), sans les « confondre » avec cette environnement, c'est-à-dire sans les penser vraiment là. Cette différence entre l'environnement qui entoure l'immersion (salon où l'on lit, salle de théâtre ou de cinéma, et ainsi de suite) et ce qui arrive dans l'immersion *n'est pas* celle qui est en question ici (alors que l'on peut en rendre compte dans ce cadre)¹⁰. Car, tout simplement, elle n'est pas une compétence spécifique à la fiction, mais plutôt une compétence nécessaire pour une lecture quelconque (où ce qui est décrit n'arrive pas simultanément sous les yeux du lecteur). Le *hors sujet* du fait que c'est de la fiction pendant les moments d'immersion fictionnel est tout à fait compatible avec une « conscience » (ou une « présence ») du fait que les événements racontés n'ont pas lieu dans l'environnement immédiat du lecteur.

3.1.2 Les mises en jeu du fait de fiction

Où l'on considère quelques cas où le fait de fiction entre en jeu de manière focalisée. En employant le cadre proposé dans le

⁸Schaeffer (1999, p182).

⁹C'est l'exemple de Schaeffer, *ibid.* Alors qu'il discute un exemple relatif à l'émetteur d'un texte de fiction (l'auteur), ses conclusions sont censées s'appliquer également au cas du récepteur (le lecteur).

¹⁰Par ailleurs, l'intérêt de la réalité virtuelle réside précisément dans l'écrasement de la distinction entre le monde réel qui entoure l'activité fictionnelle et l'activité réel qui correspond et oppose à l'activité fictionnelle. Lire un texte de fiction implique des différences et des ressemblances intéressantes par rapport à lire un texte non fictionnel (plutôt que par rapport à s'asseoir dans son chambre) ; regarder un film ou une pièce fictionnel implique des différences et des ressemblances intéressantes par rapport à regarder un film documentaire ou une pièce non fictionnelle (plutôt que par rapport à s'asseoir dans une salle noire) ; mais « jouer » une réalité virtuelle, ne comporte-elle pas les rapports intéressants précisément avec se comporter « ordinairement » ?

chapitre précédent, on peut rendre compte de ces mises en jeu comme l'apparition d'un objet fiction dans la structure instantanée de la conscience au moment approprié. De manière générale, les différentes modalités de sa mise en jeu correspondent aux différents aspects d'une structure instantanée qui peuvent lui correspondre et aux différentes manières d'organiser ces aspects dans la structure – plus ou moins en arrière-plan, entrant dans certains rapports plus ou moins forts avec d'autres aspects de la structure, et ainsi de suite.

Aux moments d'immersion fictionnelle, le fait de fiction est *absent* de la structure instantanée. À l'inverse, le fait de fiction est *présent* dans la structure instantanée, *d'une manière ou d'une autre*, à des moments où le fait que le texte est fictionnel est « en jeu ». Cette présence connaît plusieurs « modalités » possibles, selon qu'il s'agisse de la présence d'un objet, d'une relation ou d'une configuration de la structure, ou, pour reprendre la terminologie de §2.2.2, selon que le fait de fiction soit *focalisé* ou non. Dans ce paragraphe, on considère la présence *focalisée* du fait de fiction à un moment, laquelle connaît plusieurs variations selon les autres objets et aspects de la structure instantanée. Il s'agit donc des structures instantanées appropriées à certaines situations où le sujet « enjambe » la frontière entre fiction et réalité, ayant les deux dans la vue de l'esprit en même temps, pour ainsi dire, telles des situations où le sujet considère la fiction « du dehors » ou des situations où le fait qu'il s'agit de la fiction entre en jeu à un moment d'immersion fictionnelle.

On commence avec l'exemple d'un moment où le sujet est focalisé sur le fait que Holmes soit un personnage fictionnel. On a vu auparavant que le fait d'avoir quelque chose *in focus* ou de lui prêter de l'attention à un moment se rapporte à un objet (ou des objets) dans la structure instantanée de la conscience à ce moment (§2.2.2) ; de même, le fait d'être focalisé sur le fait que quelque chose soit fictionnel correspond à un objet (ou des objets) dans la structure. Pour le dire grossièrement et simplement, le fait de tenir compte du fait que Holmes soit fictionnel correspond à l'apparition d'un objet *fiction* (ou même *Aventures de Sherlock Holmes*¹¹) relié (par une relation « ... vient de ... ») à l'objet *Holmes* dans la structure instantanée de la conscience à l'instant approprié.

Cet exemple simple sert seulement à *indiquer* le *genre* d'aspect de la structure instantanée de la conscience qui correspond à la mise en jeu de la fiction. La théorie du mental développée dans Chapitre 2 permet de rendre compte des multiples « manières » ou « degrés » dont le fait de fiction pourrait être en jeu

¹¹Évidemment, à des moments où il s'agit des personnages provenant des œuvres différentes, il y aura des objets différents pour chaque texte dans la structure instantanée de la conscience.

en termes des différentes modalités de son apparition dans une structure instantanée. Comme le montre l'exemple ci-dessus, si le fait de fiction est focalisé, il y a un objet *fiction* dans la structure instantanée ; de plus, cet objet est plus ou moins à l'arrière-plan selon que le fait de fiction l'est. Comme on l'a vu dans le paragraphe précédent, si le fait de fiction est *hors sujet*, il n'y a ni objet ni relation dans la structure instantanée qui s'y rapporte. Comme on le verra dans le paragraphe suivant, le fait de fiction pourrait être non focalisé ; en ce cas il se rattache à une relation dans la structure instantanée. Pour illustrer ces possibilités sur des cas plus complexes, mais toujours impliquant l'apparition *focalisée* du fait de fiction, on considère deux exemples des entrées en jeu du fait de fiction pendant l'immersion fictionnelle.

Il s'agit d'abord de l'hésitation du lecteur au moment où l'on interrompt sa lecture en lui posant la question de l'existence de Holmes, exemple que l'on a relevé dans §§1.4.1 et 1.5.2 comme problématique pour les théories traditionnelles de la fiction. À un tel moment, la structure instantanée de la conscience comprend toujours l'objet *fiction* relié à Holmes. L'ambiguïté et l'hésitation tiennent au fait que, dans la structure, il n'y a pas de « priorité » entre le fait qu'il y a un objet Holmes et le fait que cet objet est relié à l'objet *fiction* ; c'est-à-dire, entre « l'existence » au sens « large », pour ainsi dire, et l'existence « effective ». L'hésitation se rend facilement dans ce cadre théorique comme un aspect *structurel* de la structure instantanée à cet instant : de la *configuration* des objets tels que *fiction* et Holmes. L'issue de cette hésitation, c'est-à-dire la réponse à la question, sera décidée dans la succession des structures instantanées suivantes, où l'un des aspects gagnera la priorité, alors que, conforme au rapport intime mais certainement pas déterminant entre la structure instantanée de la conscience et sa dynamique (§2.3.4), il *se peut que* la forme ou l'organisation de cette structure à cet instant correspondra à une tendance à la résolution finale dans l'un sens ou l'autre.

Enfin, pour prendre un exemple qui met en avant les affects relatifs à la fiction et donc qui a trait à la question **B.** de §1.1.1, il s'agit de l'instant où le lecteur s'efforce de se libérer des sentiments éprouvés à l'égard du texte en mettant en jeu son caractère fictionnel, instant qui a également posé problème pour les théories traditionnelles (§1.4.1). Encore une fois, alors que la question générale du rapport entre l'affect et la croyance au fait de fiction réfère à la succession dynamique dans laquelle le « conflit » entre eux se joue et trouve son issue, il n'est pas difficile, dans le cadre proposé, de comprendre *en gros* la structure instantanée de la conscience à cet instant : elle comporte un objet *fiction* relié à d'autres objets appropriés (cela rend compte du caractère focalisé du fait de

fiction) et, de plus, elle a une certaine propriété globale ou une certaine « forme » (ce qui correspond au sentiment éprouvé, §2.5.2). Le rapport entre l'objet *fiction* et la forme globale de la structure est complexe et difficile : l'exemple du lecteur qui n'arrive pas à effacer tout de suite son sentiment en rappelant le fait qu'il s'agit de la fiction montre que, au moins pour certaines formes (c'est-à-dire, sentiments), le fait que l'objet *fiction* apparaît et entre en relation avec d'autres objets de la structure ne nuit pas à la forme générale de la structure (et donc au sentiment qui lui correspond). Les questions de savoir si certaines formes sont incompatibles avec l'objet *fiction* (si certains sentiments s'absentent dès que l'on fait attention au fait de fiction), ou si certaines formes sont bousculées plus ou moins rapidement ou facilement dans les structures instantanées suivantes par la mise en jeu de la fiction (si l'on se dérobe du sentiment plus ou moins rapidement ou facilement en se rappelant que c'est fictionnel) sont des questions complexes, exigeant une considération des suites des consciences relatives à la fiction.

3.1.3 Confusions évitées

Où l'on examine des cas où le lecteur évite une possible confusion entre la fiction et la réalité, qui sont en outre des exemples typiques d'une mise en œuvre de manière non explicite du fait de fiction. Dans la mesure où le fait de fiction est en jeu mais non focalisé, elle s'associe à des relations ou à des configurations des relations dans des structures instantanées de la conscience à des moments appropriés; le fait d'éviter de telles confusions peut d'ailleurs être compris en termes des conséquences dynamiques suggérées par ces relations ou ces configurations.

Au-delà de l'immersion fictionnelle, le fait de pouvoir y entrer et en sortir, et de pouvoir « enjambrer » la fiction et la réalité de plusieurs manières, l'élément final de la compétence fictionnelle tient au fait que le lecteur de fiction ne *confond* pas fiction et réalité dans les circonstances où une telle confusion est possible. En effet, non seulement le lecteur « ordinaire » évite ou au moins peut éviter les actions et les dits qui impliquent une telle confusion mais de plus il le fait sans devoir se dire ou se figurer de manière *explicite* ou *délibérée* que ceci est fictionnel et cela ne l'est pas. Cette compétence s'explique donc le plus naturellement en termes de « capacités » ou de « tendances » du lecteur plutôt que d'une attention lourde au caractère fictionnel de tel ou tel événement ou personnage. Dans l'optique de la théorie du mental développée ci-dessus (notamment §2.5.3), une capacité ou une tendance, dont une tendance que l'on associe à une croyance ou,

en l'occurrence, à la connaissance que le texte est fictionnel, « s'actualise » ou « s'exerce » dans des *successions* des structures instantanées de la conscience. En outre, une croyance qui est « en jeu » ou qui exerce une influence mais qui n'est pas « en relief » ou « explicite » dans l'attention, peut être associée à une relation dans la structure instantanée de la conscience à un instant particulier de la succession appropriée, ou, plus généralement à une configuration de relations de la structure. De la même façon, si le fait de fiction exerce une influence sans être explicite, il semble qu'il se rapporte également à une relation ou une configuration de relations d'une structure instantanée à un instant particulier. Le fait de fiction est présent dans les structures instantanées à des moments où des confusions potentielles sont évitées, mais de manière *non focalisée*.

Les cas pertinents pour mettre en évidence cette « apparition non focalisée » du fait de fiction dans une structure instantanée sont ceux où les événements ou les personnages d'une fiction risquent « d'infecter » les croyances ordinaires du lecteur, et par conséquent ses comportements ordinaires, et *non pas* ceux où les événements et les personnages décrits dans le livre s'opposent à l'environnement de la lecture (§3.1.1). La différence entre la lecture de la *Noé* et celle de l'*Histoire de la Révolution Française*, entre regarder un film fictionnel ou un film documentaire, est que dans les premiers cas on ne devrait pas laisser mêler les événements et les personnages du texte parmi ceux que l'on pense être réels, alors que dans les derniers cas, on devrait. Il faut donc des exemples tel celui de l'amateur de Sherlock Holmes qui, au cours de sa vie quotidienne, se trouve face aux titres de propriété de 221b Baker Street datant de la fin du dix-neuvième. Alors que tout probablement la figure de Holmes « vient à l'esprit », il ne s'étonne pas de ne pas voir le nom de Sherlock Holmes dans les titres de propriété. D'autres exemples pertinents sont ceux qui sont relatifs à des contextes fictionnels, tel celui du lecteur qui comprend au début du roman que l'histoire a lieu à Paris, mais apprend ensuite qu'il y a des appartements en haut de la Tour Eiffel¹² : la différence avec le cas réel lui vient tout probablement à l'esprit, mais il ne lui faut pas souligner le point que c'est la fiction pour pouvoir continuer sa lecture en tranquillité.

Pour analyser ces cas, il *ne faut pas* qu'il y ait un objet spécifique *fiction* relié à l'objet « fictionnel » (Holmes, la Tour Eiffel habitée) *ni* un objet *réel* relié aux objets « réels » (vrai propriétaire de 221b Baker Street, Tour Eiffel peuplée de boutiques touristiques) dans la structure instantanée de la conscience à un instant où la confusion menaçante est toujours possible mais où le risque est en train d'être désamorcé. Au contraire, la présence de tels objets correspondrait

¹²Pavel (1986).

aux cas où le lecteur *focalise* explicitement sur le fait que l'un est fictionnel et que l'autre ne l'est pas, alors que, comme on l'a déjà souligné, le lecteur *ne fait pas entrer* de manière *délibérée* le fait qu'il s'agit de la fiction lorsqu'il évite ces confusions éventuelles, mais il les évite de manière tout à fait naturelle. De fait, ceci correspond à une configuration des relations de la structure instantanée selon laquelle l'un des objets (Holmes ou le vrai propriétaire, Tour Eiffel habitée ou Tour Eiffel touristique) entre dans des relations intéressantes et pertinentes avec d'autres objets de la structure et l'autre n'y entre pas : configuration où le vrai propriétaire de 221b Baker Street est relié à d'autres objets pertinents pour les activités actuelles et réelles du sujet (par exemple, noter le nom)¹³ ; configuration où la Tour Eiffel habitée entre dans les relations pertinentes avec des objets pertinents pour la lecture actuelle (par exemple, les personnages, les événements). C'est le degré « d'isolation » de l'objet « étranger » (Holmes dans la réalité, la Tour Eiffel touristique dans la fiction) par rapport aux autres objets qui correspond au fait de mettre en œuvre, qu'il soit de manière non explicite et non focalisée, son savoir que c'est de la fiction plutôt que de la réalité (ou vice versa). C'est une « apparition » importante de la fiction en tant que *relation* ou *configuration de relations* de la structure instantanée de la conscience à cet instant, laquelle correspond, dans les successions appropriées, à l'évitement naturel et non délibéré d'une potentielle confusion entre réalité et fiction.

3.2 Chevauchement entre fiction et réalité

Les deux questions de §1.1.1 qui ont été mises sous le nom de « chevauchement » entre fiction et réalité ont trait, d'une part, aux ajouts des aspects mentaux « réels » ou ordinaires (croyances, connaissances, préjugés) aux propos du texte pendant, et éventuellement après, la lecture de la fiction (question **C.**) et, d'autre part, à l'infiltration éventuelle des personnages, notions et conceptions introduits dans la fiction dans le « domaine mental » consacré à la réalité (question **D.**). La question est de comprendre exactement *quels* aspects réels et *quels* aspects fictionnels entrent dans *quels* rapports (§1.1.3), ce qui requiert, on a vu dans les chapitres précédents, un recours à la dynamique « crue » de la conscience. En l'occurrence, il faut non seulement comprendre les aspects et les propriétés des « conduites » du sujet qui ont affaire à la fiction, mais aussi bien des aspects de ses « conduites ordinaires » dans des situations « réelles », c'est-à-dire, pour l'exprimer dans le vocabulaire du cadre proposé, comprendre

¹³On rappelle le sens large d'objet qui est employé ici (§2.1.3).

et bien les aspects et les propriétés des successions des structures instantanées de la conscience à des moments où la fiction est en jeu, *et bien* les aspects et les propriétés des successions des structures aux moments de la vie ordinaire, *et bien* les rapports entre les deux. Si l'on prend ce dernier rapport dans l'un sens, on repère des croyances, tendances, et préjugées qui contribuent à ou même étaient la lecture et l'appréciation du texte et qui ont des conséquences sur les avis du lecteur à l'égard de ses propos et de sa valeur (question **C.**). Si l'on le considère dans l'autre sens, on repère des aspects, thèmes et caractéristiques glanés du texte pendant la lecture qui s'infiltrèrent dans les pensées, préjugés, actions et jugements « réels » ultérieurs (question **D.**).

Comme on l'a déjà mentionné, l'étude approfondie de la question centrale, celle de comprendre *quels* aspects de ces successions entrent dans *quels* rapports, appartient à la dynamique. Dans la mesure où l'on n'aborde pas pleinement la question de la dynamique dans cette thèse, on ne saurait pas offrir une réponse complète aux questions relatives au chevauchement entre fiction et réalité. On se contente en revanche de présenter la forme que prennent ces phénomènes de chevauchement dans le cadre de la théorie proposée du mental, et de repérer des propriétés et des conséquences pertinentes dans des structures instantanées particulières. Dans le premier paragraphe, consacré à la question **C.**, on arrive à rendre compte de certaines modalités et propriétés des irruptions des aspects réels à des moments de la lecture de la fiction. Bien entendu, cette considération ne suffira pas pour épuiser la question générale **C.**, qui a trait également aux « vérités fictionnelles » auxquelles aboutissent le lecteur à *la fin* de sa lecture, et qui appellent par conséquent une considération des rapports entre les structures de la conscience à des moments pendant la lecture et les successions des structures après la lecture où le lecteur réfléchit sur le texte, où il le retourne dans sa tête, et où il affirme des avis et des jugements à son égard. Dans le deuxième paragraphe, en guise d'une considération de la question **D.**, on considérera trois exemples de « transferts » des aspects provenant de la fiction dans la réalité, qui permettent, quoique à un moindre degré, une approche en termes des structures instantanées particulières.

3.2.1 Question C. : L'irruption de la réalité

Où l'on analyse, à l'aide des concepts développés dans le chapitre précédent, des moments particuliers où les connaissances et plus généralement les attitudes du lecteur à l'égard de la réalité interviennent dans sa lecture d'un texte de fiction. Un aspect réel peut

entrer en jeu tantôt de manière focalisée tantôt de manière non focalisée : dans ce dernier cas, l'aspect réel exerce une influence sur la lecture sans qu'il soit explicitement constaté par le lecteur. C'est en termes de ces genres d'influences, qui s'exercent pour l'essentiel dans la dynamique, que l'on aborde les questions de la « concrétisation » des événements ou personnages dans la lecture, ou des « vérités fictionnelles » auxquelles le lecteur aboutit. Bien que l'on n'entre pas dans les détails de ces questions ici, on arrive, en mettant en œuvre quelques propriétés des structures instantanées de la conscience, à expliquer certains phénomènes relatifs à la concrétisation.

Malgré le fait que la question de l'irruption des croyances « usuelles » ou « réelles » pendant la lecture d'un texte de fiction est complexe, renvoyant, dans ses détails, à la dynamique des structures instantanées de la conscience, les concepts développés dans le Chapitre 2 permettent une explication plus approfondie de certaines modalités de l'irruption de la réalité dans la fiction, de certains de ses enjeux, et de certaines de ses conséquences. Tout d'abord, en appliquant la notion d'apparition focalisée ou non des aspects dans une structure instantanée de la conscience aux croyances usuelles qui interviennent dans la lecture, on arrive à une distinction grossière entre deux manières pour une croyance d'entrer en jeu – de manière focalisée ou non. Les croyances qui entrent en jeu de manière non focalisée ne correspondent pas à des objets dans des structures instantanées appropriées, mais seulement à des relations ou des configurations de relations des structures instantanées ; comme telles, le sujet ne se rend pas explicitement compte de leur irruption dans sa lecture. Néanmoins, conformément aux conséquences de la continuité de la dynamique discutées dans §2.3.4, ces croyances ont toujours une influence, voire une influence importante, sur la dynamique de la lecture. Enfin, l'entrée en jeu des croyances, connaissances ou préjugés « réels » ou « ordinaires » du lecteur aboutit à l'établissement d'un certain nombre de « vérités fictionnelles » qui débordent les propos du texte, ou autrement dit une certaine « concrétisation » des propos nus du texte. La localité des structures instantanées de la conscience permet de rendre compte du caractère partiel de cette concrétisation – seulement certains aspects sont concrétisés – et du « manque de conscience » de la part du lecteur du grand nombre de questions qui sont laissées ouvertes par le texte sans être concrétisées dans la lecture.

Il y a deux manières pour un aspect, une croyance, une connaissance, une entité « réelle » d'apparaître dans une structure instantanée de la conscience à

un instant particulier¹⁴ : *grosso modo*, en tant qu'objet ou en tant que relation. D'un côté, si une « entité réelle » (événement, phénomène *etc.*) apparaît comme un *objet* dans une structure instantanée pendant la lecture, elle est focalisée. Par exemple, l'entité réelle *Londres* figure comme objet dans les structures instantanées à des moments de la lecture des *Aventures de Sherlock Holmes*¹⁵, notamment à des moments où elle est *explicitement* mentionnée. Le fait pour une entité (phénomène, événement, aspect) réelle d'être focalisée à un instant *n'implique pas* cependant que *le fait qu'elle soit réelle plutôt que fictionnelle* est lui-même focalisé. Au contraire, en pleine immersion fictionnelle, les structures instantanées comportant l'objet *Londres* ne comportent ni un objet *réel* relié à l'objet *Londres* (à la façon de l'objet *fiction* de §3.1.2) ni une relation ou une configuration qui met en jeu le fait que *Londres* soit réelle plutôt que fictionnelle (à la façon de la relation associée au fait de fiction dans §3.1.3). Comme on l'a vu dans §3.1.1, dans ces cas d'immersion, le fait qu'il s'agit de fiction – et par conséquent la *distinction* entre fiction et réalité – manque *dans* la structure instantanée, à telle enseigne que le caractère réel de *Londres* est *hors sujet* à cet instant.

De l'autre côté, les *relations* dans les structures instantanées à des moments de lecture « proviennent » souvent de la réalité. Des inférences opérées dans la fiction fournissent des exemples particulièrement clairs de ces genres de relation. Pour reprendre l'exemple du lecteur qui lit que le héros risque de tomber d'un escarpement élevé (§1.1.1) : pour tirer l'inférence que le héros est dans une situation dangereuse, avec toutes les conséquences affectives qui s'ensuivent, il faut un « ajout » d'une connaissance relative à un fait réel, en l'occurrence, grossièrement, le fait qu'un être humain qui tombe d'un escarpement élevé se fait du mal. Le fait pour le lecteur de tirer cette inférence apparaît d'abord comme une propriété de la succession des structures instantanées de sa conscience, à savoir le fait de progresser d'une structure instantanée comportant des propos du passage du texte à une structure instantanée comportant (de surcroît) le fait que le héros est en danger. Or, d'après le rapport étroit entre l'état instantané et la dynamique (§2.3), et plus spécifiquement entre des propriétés des successions des structures instantanées et des aspects dans les structures instantanées

¹⁴ *Stricto sensu*, les croyances ou entités réelles ne figurent pas *dans* les structures instantanées, mais plutôt il y a des aspects de ces structures qui « correspondent » ou sont associées aux croyances ou entités réelles ; voir les remarques de §§2.2.1, 2.4 et §1.3.3 sur ce point. On rappelle par ailleurs que le sens d'apparaître « dans » un état instantané de la conscience, et donc des « entités » en question, est, conformément à l'optique de ce travail, « internaliste » ; voir §§1.1.2 et 2.1.1. On omet ces précisions dans la suite de la discussion.

¹⁵ On rappelle le sens large d'objet (§2.1.3) qui sera en vigueur pendant toute cette discussion, et le fait qu'il se réfère seulement aux aspects de la psychologie du lecteur.

particulières, cette propriété peut être associée à une relation dans une structure instantanée à un instant particulier, en l'occurrence une sorte de relation « d'inférence » entre les propos du passage et le danger. De même que, dans les cas d'entrée en jeu d'une croyance discutés dans §2.5.3, on associe la croyance à la relation appropriée dans la structure instantanée appropriée, de même cette relation « d'inférence » est « associée » à ou conçue comme « recélant » la croyance *usuelle* qui soutient l'inférence, en l'occurrence la croyance que des êtres humains ne survivent pas à de telles chutes. C'est donc un exemple d'une relation dans une structure instantanée particulière qui est associée à une croyance *usuelle*.

Le fait que des éléments du « corpus ordinaire » peuvent s'immiscer dans la lecture en tant que relations des structures instantanées est d'une importance particulière. En effet, bien qu'un certain nombre de relations dans une structure instantanée à un moment d'immersion soient « fictionnelles », dans le sens où les « faits » qui leur correspondent se réfèrent au texte de fiction, la plupart se rattachent apparemment aux connaissances, attitudes, préjugés, suppositions, inclinations du lecteur relatives à la réalité. Il y a non seulement les connaissances qui permettent des déductions plus ou moins sûres (telle l'éventuelle chute de l'escarpement), mais également les suppositions et les inférences plus précaires et passagères, qui sont normalement valides dans la réalité (par exemple, lorsqu'on dit, ordinairement ou dans la fiction, qu'ils sont mariés, il est entendu qu'ils sont mariés l'un avec l'autre). Si l'on peut dire, la lecture de la fiction se déroule dans un « bain » de réalité¹⁶.

En outre, puisque ce « bain » est constituée de relations, les aspects « réels » qui interviennent dans la lecture et qui agissent sur les aspects fictionnels sont *non focalisés* (§2.2.2)¹⁷. Autrement dit, le lecteur ne tient pas explicitement compte de leur entrée en jeu. L'irruption de ces aspects, inclinations, connaissances « réels » ou « usuels » dans la lecture est donc moins apparente, et souvent difficile à repérer. En particulier, pour certains aspects des entités ou des événements fictionnels, il pourrait ne pas être clair de distinguer entre la partie qui renvoie aux passages du texte, et la partie qui a ses origines dans des associations, suppositions, inférences, jugements que le lecteur fait normalement dans la réalité. Conformément au rapport entre la structure instantanée de la conscience et sa dynamique, ceci vaut aussi bien pour une structure instantanée de la conscience à un instant particulier (difficulté d'identifier des relations avec des aspects réels ou fictionnels), que pour la dynamique de la succession de

¹⁶Voir sur ce point la discussion du rapport entre le jeu et la réalité dans §1.2.2.

¹⁷C'est-à-dire, les aspects mêmes et non (seulement) le fait qu'ils sont réels, sont non focalisés.

ces structures qui constitue la démarche de la lecture (difficulté d'identifier les propriétés des successions des structures avec des aspects du « corpus usuel » plutôt que « fictionnel » des attitudes du lecteur).

Les effets et les influences clandestins des croyances, connaissances, préjugés « usuels » est bien connu et peu compris. En deux mots, le bon auteur peut invoquer des images, des aspects, des préjugés, des oppositions, des ressemblances, de telle manière que l'image que se fait le lecteur des événements et des personnages décrits déborde largement les propos « crus » du texte. La question de cet « étoffement » du texte par « ajout » d'attitudes « réelles » a reçu plusieurs noms : elle est la question des « vérités fictionnelles » – les énoncés « vrais selon le texte » qui ne se limitent point aux simples affirmations contenues entre les couvertures¹⁸ ; également, elle appartient à ce qu'Ingarden appelle la « *concrétisation* », par laquelle sont « remplis » les « lieux d'indétermination » que, selon lui, comporte le texte¹⁹.

La question de savoir comment un texte est étoffé, quels aspects « réels » y sont ajoutés dans la lecture, quelles « vérités fictionnelles » le lecteur en tire, quelles « concrétisations » des personnages et des événements il opère, est une question relative aux successions de structures instantanées de la conscience de la lecture et la considération du texte, et aux propriétés de ces successions ; elle appartient à la dynamique. Dans ce travail, on n'aura rien de plus précis à dire à ce sujet. Néanmoins, la théorie du mental et notamment de la structure instantanée de la conscience développée dans le chapitre précédent offre une compréhension approfondie de certaines propriétés relatives aux entrées en jeu des aspects réels à des instants particuliers de la lecture²⁰. À savoir, premièrement, le fait que la concrétisation opérée à un moment de lecture n'est que *partielle* et, deuxièmement, le fait que les « lieux d'indétermination » qui restent *ne sont que rarement constatés* par le lecteur.

Pour illustrer ces points, on emprunte à Ingarden l'exemple de la description d'un vieillard qui, même si elle n'affirme pas explicitement qu'il a des cheveux gris, pourrait bien inciter dans « l'esprit du lecteur » l'image d'un homme

¹⁸On reconnaît normalement l'article de référence à Lewis (1978). Conformément au contexte « internaliste » de cette discussion, on a affaire ici aux vérités fictionnelles *selon l'avis du lecteur*. On aura l'occasion de revenir sur la question de la vérité fictionnelle dans l'Introduction à la seconde partie et dans §4.1.2.

¹⁹Ingarden (1965, §38 et Ch. 13) ; également, 1968, §11.

²⁰Cette discussion porte sur les moments particuliers *de la lecture*, et donc sur les avis et les images du lecteur qui sont concomitantes à sa lecture. Il est une question légèrement différente de savoir les avis et les images que le lecteur se fait *à la fin* de la lecture – les concrétisations et les vérités fictionnelles auxquelles il est finalement parvenu, pour ainsi dire.

à cheveux gris²¹. Cette description n'incite cependant pas une précision dans « l'esprit du lecteur » du nombre de cheveux sur la tête²². Deux aspects de la concrétisation ressortent de cet exemple : premièrement, elle est souvent seulement partielle (le nombre de cheveux n'est pas concrétisé) ; deuxièmement, souvent le lecteur ne remarque pas les « lieux d'indétermination » qui restent (il ne remarque pas qu'il ne connaît pas le nombre de cheveux)²³. La *localité* de la structure instantanée de la conscience permet de comprendre ces phénomènes. Tout d'abord, puisque seulement certains aspects ou propriétés éventuels sont plus ou moins *en jeu* ou *présents* dans une structure instantanée à un instant particulier, seulement certaines propriétés éventuelles du personnage sont « concrétisées » à un instant : alors que, par exemple, l'objet *cheveux gris* pourrait être plus ou moins à l'arrière-plan dans la structure instantanée à l'instant approprié, l'objet *nombre de cheveux* serait tout probablement totalement absent. Ensuite, les propriétés *absentes* de la structure instantanée qui demeurent indéterminées ne sont pas, par le fait même d'être *hors sujet*, *remarquées* par le lecteur à ce moment de sa lecture. Bien évidemment, il n'est pas le cas que tous les lieux d'indétermination non remplis ne sont pas constatés par le lecteur : au contraire, comme l'a fait souligner Iser, il est parfois important pour le lecteur de remarquer certains « trous » ou « blancs » du texte²⁴. Ces cas des lieux d'indétermination constatés peuvent être également compris dans le cadre proposé, en termes de l'apparition de l'objet approprié (par exemple *couleur de cheveux* dans un cas où le lecteur constate que la couleur de cheveux n'est pas spécifiée) en tant que relié à un objet tel *non spécifié*. De toute manière, il reste que la localité de la structure instantanée permet de comprendre la différence entre les propriétés concrétisées et les propriétés non déterminées mais constatées comme telles d'une part, et les propriétés non déterminées mais ignorées d'autre part.

En résumé, l'irruption de la réalité lors de la lecture de la fiction se comprend à des moments particuliers en termes des aspects (objets, relations) de la structure instantanée à ces moments. Les propriétés des structures instantanées permettent de comprendre le genre de complétude que l'on éprouve et le genre de concrétisation « automatique » que l'on effectue à des moments de

²¹Ingarden (1968, p392).

²²Lewis (1978, p270).

²³Ingarden (1965, p213) : « on ne remarque à la lecture aucune « lacune », aucun « lieu d'indétermination », dans l'appréhension esthétique » ; ou 1968, p53 : « [les objets décrits] comportent toujours les lieux d'indétermination, mais ce qui est déterminé est pour ainsi dire tourné du côté du lecteur, et couvre les trous dans la détermination. »

²⁴Iser (1976, surtout Ch. 8).

lecture. Bien entendu, il faut toujours comprendre *quels* aspects sont concrétisés ou non, *quels* aspect réels interviennent et quand, *quelles* influences ces aspects supplémentaires ont sur la suite de la lecture (ou du raisonnement à l'égard du texte) : ce sont des questions pour la dynamique.

3.2.2 Question D. : L'infiltration de la fiction

Où l'on considère la question de l'infiltration de la fiction dans la réalité à travers trois cas de figure. Il s'agit, d'abord, des énoncés qui font référence à des personnages fictionnels dans des contextes qui ne peuvent qu'être non fictionnels, tel « Holmes est un personnage fictionnel » ; ensuite, des emplois explicites des aspects relevant d'un texte de fiction pour décrire quelque chose de réel, tel « untel est un 'Holmes' » ; et, enfin, des influences plus générales de la lecture d'un texte de fiction sur la conception qu'a le lecteur de la réalité, tels les effets relatifs à sa moralité. Bien que les difficultés que posent ces cas renvoient en partie à la dynamique des structures instantanées de la conscience, on arrive à comprendre des aspects importants, surtout des deux premiers cas, en considérant les structures instantanées individuelles à des moments appropriés.

La question **D.** de §1.1.1 a trait aux infiltrations des aspects fictionnels sur le terrain de la réalité. Comme ailleurs, une réponse complète à cette question exigerait une compréhension de la dynamique des successions des structures instantanées de la conscience impliquées aussi bien dans la lecture de la fiction qu'aux moments d'usage, de mise en jeu ou de la manifestation de certains aspects, concepts, personnages ou thèmes provenant d'un texte de fiction dans un contexte réel. Il ne s'agit pas ici d'aborder la question en détail, mais seulement de considérer quelques aspects des trois sortes « d'infiltration » de la fiction dans la réalité remarquées dans §1.1.3, en tant qu'elles se présentent à la lumière de la théorie du mental proposée. À savoir, il s'agit, premièrement, de la « référence » à un personnage ou aspect fictionnel dans un contexte qui ne pourrait être que non fictionnel (cas typique : « Holmes est un personnage fictionnel ») ; deuxièmement, des *emplois* explicites des aspects, concepts ou personnages plus ou moins fictionnels dans la discussions de la réalité (cas typique : « untel est un 'Holmes' ») ; et troisièmement, des *effets* ou *influences* de la lecture de la fiction sur la conception de la réalité du lecteur (cas typique : les éventuels effets de *Tristram Shandy* de Lawrance Sterne, œuvre polémique contre l'empirisme lockéen, sur la vision que le lecteur a sur le monde). Le premier cas est discuté ou mentionné par des philosophes comme Lewis, Wal-

ton, Currie²⁵ ; le deuxième est l'exemple typique discuté par Goodman²⁶ ; le troisième cas est thématique par Iser ou par Ricoeur²⁷. Il s'avère que, alors qu'il y a des choses à dire au sujet des deux premiers cas dans des structures instantanées particulières, le troisième cas ne pourrait être traité que du point de vue « dynamique ».

Un exemple du premier type d'infiltration est un moment où le sujet affirme ou considère ou « a en tête » le fait que Holmes est un personnage fictionnel. L'attitude du sujet par rapport à ce fait ne pourrait pas être celle de *faire comme si*, puisqu'il n'est pas une vérité fictionnelle que Holmes est fictionnel ; pourtant, il diffère d'une croyance « usuelle » dans la mesure où elle fait référence à quelque chose que le lecteur sait ne pas exister, à savoir « l'entité fictionnelle » Holmes. Cette tension relève plutôt de la simplicité de la distinction entre fictionnelle et réelle (ou usuelle) que l'on tente d'appliquer aux attitudes du lecteur que d'un vrai problème ; en effet, dans la structure instantanée de la conscience à cet instant, l'état de choses est assez clair. L'objet *Holmes* n'est pas un objet de statut radicalement différent d'autres objets, mais acquiert une spécificité seulement par le fait qu'il est relié à un objet *fiction*, par une relation comme « vient de » (§3.1.2) ; c'est d'ailleurs en cette configuration d'objets que consiste la pensée que Holmes est fictionnels. De manière générale, le fait que les « entités fictionnelles » ou les « événements fictionnels » sont des objets (*a priori* « ordinaires ») des structures instantanées où ils sont *focalisés* permet de rendre compte des structures instantanées aux moments où le sujet affirme ou considère des « énoncés transfictionnels »²⁸, c'est-à-dire des énoncés *X* qui figurent des termes qui font référence aux entités ou événements fictionnels (et donc, « réellement », des termes à référence vide), mais tel que l'énoncé « selon le texte fictionnel, *X* » est faux (sinon dénué de sens). Par exemple, l'énoncé « Holmes aurait battu Lénine aux échecs » figure un terme (Holmes) sans dénotation réel, mais fait référence à une entité réelle qui n'apparaît pas dans les œuvres de Conan Doyle. Dans la structure instantanée à un moment où le sujet considère cet énoncé, Holmes et Lénine figurent comme objets, avec une relation appropriée entre eux. Par ailleurs l'objet *fiction*, et par conséquent le fait que Holmes soit fictionnel, pourrait bien ne pas apparaître dans cette structure – il y serait *hors sujet*. Cette possibilité d'apparition des « entités fictionnelles » dans des questionnements « réels », sans chaperon pour rappeler

²⁵Lewis (1978, p264), Walton (1990, Partie 4), Currie (1990, §4.11).

²⁶Goodman (1978, Ch VI, §5).

²⁷Iser (1976, Ch 3) (dont est pris l'exemple de *Tristram Shandy*), Ricoeur (1983, 1^{re} Partie, Ch 3).

²⁸Pour employer le terme de Currie, *ibid.*

le fait qu'elles sont fictionnelles, rejoint l'expérience ; en effet, c'est un des sens possibles de la célèbre « apparence de réalité » des entités et des événements fictionnels, laquelle a été évoquée par ailleurs dans la première lecture proposée, dans §1.1.3, pour la question **A.** de §1.1.1. Dans §3.1.1, on a expliqué ce genre de phénomène en termes de l'absence du fait de fiction dans les structures instantanées de la conscience aux moments appropriés. Ici encore, l'objet *Holmes* n'a aucune spécificité intrinsèque qui le distingue des objets « ordinaires », si bien que, lorsque le fait de fiction est absent, il prend à fond une « apparence de réalité ».

D'un autre côté, le problème relatif à l'application des « termes fictionnels », tels « Sherlock Holmes » ou « Don Quichotte » (ou mieux, « donquichottesque »), aux « entités » (situations, individus) réelles, diffère de celui que pose les « énoncés transfictionnels ». Il s'agit non tant d'accorder une valeur de vérité littérale à un énoncé comprenant un terme non dénotant, tel « untel est un Don Quichotte », mais plutôt de comprendre le « sens métaphorique » de la vérité en jeu : comme le remarque Goodman²⁹, « « Don Quichotte » s'applique littéralement à personne mais s'applique figurativement à beaucoup d'entre nous. » Dans le présent cadre, il s'agit de comprendre les structures instantanées de la conscience où un personnage réel est mis en relation avec un personnage fictionnel dans cette façon « figurative ». Comme ailleurs, il y a deux questions reliées : premièrement, celle de l'état de la conscience à un moment de considération ou d'affirmation d'une telle « prédication fictionnelle des êtres réels » ; deuxièmement, celle des propriétés et des aspects de la dynamique de la conscience qui mènent à ou qui suivent l'affirmation ou l'acceptation d'une telle prédication. Les concepts développés dans le Chapitre 2 permettent une réponse à la première question : la structure instantanée à l'instant approprié comporte l'objet *Don Quichotte* (ou l'objet *donquichottesque*³⁰), un objet correspondant à la personne réelle, et une relation « est un » entre eux. Les propriétés qui *soutiennent* ce rapprochement pourraient être plus ou moins claires à ce moment, en fonction des autres aspects de Don Quichotte et de la personne réelle qui sont *présents* dans cette structure instantanée (en tant qu'objets ou relations plus ou moins à l'arrière-plan). En revanche, pour répondre à la deuxième question, et par conséquent aux questions de *comment* on arrive à faire une telle prédication et de *quels aspects* lui ont « donné lieu », il s'agit de repérer des propriétés détaillées des successions qui comportent de telles structures instantanées. Cette réponse appartient à une étude approfondie de la dynamique

²⁹ *Op cit.*, p103.

³⁰ Voir §2.4 sur la question de l'interprétation des objets des structures instantanées.

des structures instantanées, et, comme telle, à un stade ultérieur de la théorie ébauchée dans le chapitre précédent.

En dernier lieu, les théoriciens tels Iser et Ricoeur mettent l'accent sur la possibilité dans la lecture des textes de fiction de « contrer les problèmes produits par le système [de pensée] » ou de « refigurer un monde »³¹. Alors que ces philosophes tendent à concentrer sur les textes littéraires, il s'agit ici, comme on l'a souligné dans l'Introduction et dans §1.1.3, de la fiction en générale – aussi bien les œuvres les plus reconnues que la fiction *bon publique* – de telle manière que ce genre de « refiguration » représente plutôt une *possibilité* des textes de fiction et de leur lecture qu'une *propriété nécessaire*. En vérité, la question de ces « effets » possibles de la lecture sur la conception de la réalité renvoie à la dynamique des successions de structures instantanées : puisque les croyances correspondent d'abord aux propriétés des successions des structures instantanées de la conscience (§2.5.3), la question des *changements* des croyances suite à la lecture d'un texte de fiction est effectivement celle des *changements* dans les propriétés de certaines successions de structures instantanées avant et après la lecture³². Elle est ainsi une question pour la dynamique, et d'ailleurs une question particulièrement épineuse dans la mesure où, à la différence des exemples d'applications métaphoriques des termes fictionnels aux entités réelles discutés par Goodman, il n'est pas toujours facile de trouver un moment spécifique où ces « effets » ou ces « changements de croyance » apparaissent explicitement dans la structure instantanée de la conscience, d'autant plus qu'elles peuvent s'appliquer à des croyances « structurelles » (les « systèmes de pensée », comme dit Iser), qui sont rarement focalisées.

En somme, les questions épineuses du « chevauchement » entre fiction et réalité renvoient pour l'essentiel à la dynamique des successions des structures instantanées de la conscience. On pourrait facilement comprendre les moments d'irruption des « faits réels » dans la lecture, aussi bien que des infiltrations des entités ou des aspects fictionnels dans un contexte « réel », en termes des aspects des structures instantanées à ces moments : la notion de structure instantanée développée antérieurement permet d'aborder la situation à ces moments. Pourtant, d'autres questions épineuses à la dynamique : quelles propriétés de succes-

³¹Iser, *op cit.*, p73 ; Ricoeur, *op cit.*, 1^{re}Partie, Ch 3., notamment, §3. Ce genre de thèse a, bien entendu, une histoire notable, sur laquelle on ne s'attarde pas ici.

³²On constate que ce que Iser discute sous le nom de « système de pensée » (*op cit.*, p70 *sq.*) ne réfère qu'à des croyances, des préjugés, des pensées, des tendances, des comportements, des attitudes du lecteur : bref, à des propriétés des successions des structures instantanées. Par conséquent, la question des changements des systèmes de pensée se ramène également à une question relative aux propriétés des successions des structures instantanées.

sions et quelles structures précédentes ont mené à une structure instantanée où ce fait réel est en jeu ? Quelles structures suivent celle-ci et quelles propriétés des successions ultérieures y aura-t-il ? L'accent sur l'importance de la dynamique est un *leitmotiv* de ce travail ; il n'est cependant nullement affirmé que la question de la dynamique ne soit facile, et on ne prétend aucunement l'aborder en détail ici.

3.3 Autres questions : États Mentaux spécifiques à la fiction

La considération de la question posée dans §1.1 – celle du rapport détaillé entre la fiction et la réalité dans la psychologie du lecteur – s'achève ici. On comprend les aspects instantanés ; pour en faire plus, il faut aborder en détail la dynamique de la structure instantanée de la conscience, laquelle, d'après le soin pris pour s'assurer qu'une telle dynamique soit pensée de manière *continue*, devrait être au moins partiellement possible (§2.3).

L'approche proposée ici ne prend aucun appui sur les notions employées par d'autres approches, tels notamment les supposés États Mentaux spécifiques à la fiction (*faire comme si*, « imagination » ; §§1.2.1 et 1.4) ; au contraire, elle puise ses sources dans les insuffisances de ces notions pour ce qui concerne les questions *détaillées* relatives à la fiction. Néanmoins, le cadre théorique développé dans le Chapitre 2 est non seulement capable de retrouver ces supposés États Mentaux spécifiques à la fiction, mais elle permet en outre de diagnostiquer les faiblesses, les échecs et les limites des théories qui se reposent sur eux. En l'occurrence, il s'agit des conséquences du fait que l'État Mental de *faire comme si* est un État Mental ordinaire et narratif, au sens de §§1.3.2 et 1.3.3, d'un côté, et des considérations précédentes relatives aux États Mentaux narratifs (§2.5.3), de l'autre. En effet, de même qu'une croyance (exemple typique d'État Mental narratif) est associée à des propriétés des successions des structures instantanées de la conscience (§2.5.3)³³, de même un État Mental de *faire comme si* se rapporte à des propriétés des successions des structures instantanées. Le fait pour le lecteur de *faire comme si* Holmes est un détective est associé, par exemple, à la succession qui mène de la structure instantanée où il prend connaissance (« fictionnelle ») du crime à *Baskerville Hall*, à une structure instantanée où il

³³On rappelle les remarques faites dans l'Introduction à §2.5 au sujet de la terminologie « relâchée » employée pour parler du rapport entre les États Mentaux et les successions des structures instantanées.

conclut que Holmes devrait trouver le coupable.

Cette perspective éclaire les problèmes qui ont touché aux tentatives de comprendre les détails de la psychologie du rapport entre fiction et réalité en termes de ces États Mentaux. En effet, de tels projets s'avèrent être empêchés par deux facteurs : le premier est le manque de distinction naturelle pré-donnée entre fiction et réalité *dans* les structures instantanées de la conscience ; le second, qui se pose de manière générale pour tout État Mental ordinaire (au sens de §1.3.2), est la complexité des rapports entre États Mentaux attribués dans des blocs d'attribution différents (voir §1.3.2).

En premier lieu, puisque le *faire comme si* est un État Mental narratif, il est associé non seulement à des successions des structures instantanées de la conscience, mais également à des objets et des relations dans certaines de ces structures (§§2.3.4 et 2.5.3). En conséquence, la distinction tranchante entre les États Mentaux « fictionnels » et « réels » (*faire comme si* et croyance) implique une distinction, dans un grand nombre de structures instantanées de la conscience, entre les objets ou relations « fictionnels » (associés à des *faire comme si*) et des objets ou relations « réels » (associés à des croyances). Or, il n'y a souvent *pas* de distinction *immédiatement présente* dans des structures instantanées particulières entre leurs objets « fictionnels » et « réels », et certainement pas une telle distinction entre leurs relations « fictionnelles » et « réelles ». Quant aux *objets*, on a vu dans §§3.1.1, 3.2.1 et 3.2.2 de nombreuses situations où la structure instantanée de la conscience comporte des objets que l'on qualifierait de « réels » ou de « fictionnels » sans que ce caractère « réel » ou « fictionnel » y soit *présent*, et *a fortiori focalisé*. On a même donné des raisons, dans §3.1.1, pour douter de la thèse défendue par Sartre et Ingarden d'une distinction *non focalisée* entre objets « fictionnels » et « réels », sous forme d'une relation « ... est fictionnel ». Du côté des *relations*, la situation est encore plus claire : il ne peut pas y avoir de distinction, présente dans une structure instantanée, entre les relations « fictionnelles » et « réelles », car, tout simplement, il ne peut pas y avoir de relation *dans la structure* entre une relation et un objet (dont un objet *fiction*), ni de relation (telle « ... est fictionnel ») qui s'applique aux relations (§2.1.3). Il n'y a donc pas nécessairement de distinction entre objets et relations « fictionnels » et « réels » dans une structure instantanée, *naturellement donnée* comme « apparente » au sujet, bien que, comme on l'a déjà remarqué, une telle distinction soit entraînée par la distinction tranchante les *faire comme si* et les croyances. Cette distinction doit être *introduite par le théoricien*.

Pourtant, l'emplacement de cette distinction, voire même la possibilité de la poser, ne va souvent pas de soi. Par exemple, le lecteur, en lisant qu'un per-

sonnage *X* pense ceci, anticipe que *X* va faire cela – la relation entre les deux objets « *X* pense ceci » et « *X* va faire cela », relève-elle d'une *croyance* usuelle (les gens font cela s'ils pensent ceci) ou associe-t-elle plutôt à quelque chose de l'ordre du *faire comme si* (ce personnage fictionnel *X* ferait cela s'il pensait ceci)? Il est une question complexe et difficile de distinguer, pour une attitude cognitive jouant un tel rôle dans la lecture de la fiction, si elle est croyance ou *faire comme si*. Les théories du *faire comme si* souffrent des difficultés à préciser la distinction sur laquelle leur approche repose.

Le second problème pour les États Mentaux du *faire comme si* touche aux États Mentaux ordinaires en général : les rapports entre des États Mentaux différents peuvent être fort complexes et difficiles. On a vu dans §§1.3.2 et 2.5 que les États Mentaux s'attribuent en bloc sur la base des comportements ou, dans le cas de la théorie proposée, des structures instantanées de la conscience ; en outre, pour les États Mentaux ordinaires, dont le *faire comme si*, ces bases sont *locales*. Si bien qu'il arrive que l'on attribue deux États Mentaux différents dans deux circonstances différentes sur deux bases différentes ; il s'agit alors de deux *blocs d'attribution différents*. Or, les rapports entre les différents blocs, et surtout des blocs à bases d'attribution différentes, pourraient être fort difficiles : quel est le rapport entre la croyance qu'un parapluie protège de la pluie, attribuée sur la base d'une situation où le sujet, ayant constaté la pluie, sort muni d'un parapluie, et les croyances qui lui sont attribuées deux semaines plus tard, où, dans une situation pluvieuse semblable, il sort sans parapluie ? Et avec les croyances attribuées au même sujet, où, âgé de quatre ans, il suit sa mère, agrippant son parapluie d'enfant ? Il s'agit apparemment des croyances attribuées dans des blocs différents qui, à la façon des différentes théories physiques (de l'aérodynamique, de solides et des matériaux, de la combustion) intervenant dans l'étude et le développement des différents aspects d'une voiture (sa forme, son corps, son engin), se rapportent clairement et précisément à leurs différents « sphères de pertinence » (structures instantanées dans la base d'attribution, aspects appropriés de la voiture), mais n'entretiennent que des rapports difficiles et partiels les unes avec les autres (blocs dont les bases couvrent les intervalles temporels lointains, théories des solides et de la combustion). Puisque, comme l'a été remarqué dans §1.3.2, un État Mental a du sens seulement relativement à un bloc d'attribution, il s'ensuit que les rapports entre les *différents États Mentaux* attribués dans des blocs différents pourraient être *fort difficiles*.

De fait, les rapports entre les États Mentaux usuels et les États Mentaux spécifiques à la fiction connaissent exactement ce genre de difficulté. Si deux États Mentaux sont mis en rapport de manières différentes avec différentes

structures instantanées de la conscience (donc, dans des blocs d'attribution différents), et si la succession des structures instantanées pose par elle-même des questions difficiles (telles la question de la dynamique ou la question de la distinction dans les structures instantanées entre aspects fictionnels et réels), on ne saurait pas s'étonner du rapport difficile entre les États Mentaux différents. Cette conclusion s'impose aussi relativement aux changements des États Mentaux (de croyances ou de *faire comme si*), et relativement aux effets éventuels de la fiction : dans les deux cas, il s'agit des changements dans les propriétés des successions des structures instantanées repérées dans différents blocs d'attribution (§§2.5.3 et 3.2.2). Une compréhension détaillée de ces phénomènes en termes des États Mentaux seulement semble fort difficile, sinon carrément impossible.

En somme, même si l'on surmonte la difficulté précédente concernant la distinction entre *faire comme si* et croyance, ces États Mentaux ne permettent pas d'éclairer les détails du rapport entre fiction et contrefactuels. Ils sont trop *grossiers* pour aborder la question de la fiction en détail ; on voit maintenant comment ils ressemblent à la vitesse moyenne en physique, « moyennant » un intervalle temporel étendu. Mais pour saisir le mouvement en détail, la vitesse moyenne ne suffit pas ; on a besoin de la vitesse instantanée. De même, pour penser à fond ces rapports entre propriétés des successions des structures instantanées, il faut un recours aux successions des structures instantanées elles-mêmes. C'est exactement la démarche préconisée dans cette thèse.

Le développement, dans le Chapitre 2, d'un certain nombre de concepts relatifs à la conscience et à sa dynamique, a donc permis d'amorcer une approche plus approfondie à la psychologie de la fiction. Dans §3.1, ils ont servis à expliquer certains aspects de la dite compétence fictionnelle, dont la capacité du lecteur de s'immerger dans sa lecture, mais également d'en sortir, de prendre son écart à l'égard des propos fictionnels, et de ne pas confondre la fiction et la réalité. Dans §3.2, ils ont permis d'éclairer certains aspects relatifs à l'irruption de la réalité dans la fiction, et également à l'infiltration de la fiction dans la réalité, et de repérer, au cœur de ces questions, la question de la dynamique des structures instantanées de la conscience. Enfin, dans la présente section, on comprend que, même laissant de côté la difficulté de préciser le sens de la distinction entre les *faire comme si* et les croyances, ces États Mentaux sont trop grossiers pour pouvoir rendre compte de ces phénomènes subtils et complexes.

Pour aborder les subtilités de la fiction, le recours à la conscience, à sa dy-

namique, et en particulier aux concepts impliqués dans leur analyse, dont la localité et la continuité, est non seulement fécond, mais nécessaire.

Deuxième partie

Jeux entrecroisés : fiction et le problème des contrefactuels

Introduction

La fiction entretient un rapport *complexe* et *difficile* avec la réalité. Souvent, on reconnaît deux aspects de ce rapport : d'une part, une *distinction* cruciale entre fiction et réalité ; d'autre part, une *similitude* complexe entre elles.

Le contrefactuel – *grosso modo*, un conditionnel subjonctif (ou irréel) tel « si l'allumette avait été frottée, alors elle s'enflammerait » – entretient lui aussi un rapport *complexe* et *difficile* avec la réalité. Dans la réalité, l'antécédent du contrefactuel est faux (l'allumette n'a pas été frotté) ; on en conclut souvent que le contrefactuel a trait non pas à la réalité, mais à une situation irréelle. Voilà une *distinction* stricte entre contrefactuel et réalité. Pourtant, la vérité du contrefactuel dépend de la réalité, en l'occurrence des propriétés réelles de l'allumette (elle est sèche, elle est bien fait) et des lois de la science (tant de friction produit tant de chaleur, et si telle substance est à tel température, alors elle s'enflamme). Si l'on insiste pour penser que le contrefactuel renvoie à une situation irréelle, ces aspects qui valent dans la réalité doivent valoir également dans la situation irréelle. En ce sens, la situation est *similaire* à la réalité. Si les contrefactuels se distinguent de la réalité, ils y préservent toujours une certaine proximité.

Distinction stricte, similitude complexe, en matière de leur relation avec la réalité, la fiction et le contrefactuel affichent un parallélisme frappant. Ce parallélisme a été d'ailleurs reconnu, plus ou moins explicitement, par divers auteurs, l'exemple le plus influent étant sans doute l'analyse que D. Lewis propose de la vérité selon un texte de fiction en s'appuyant sur sa théorie de contrefactuels³⁴. Un contrefactuel a bien affaire, selon Lewis, à des situations irréelles, donc différentes de la réalité, qui « ressemblent » néanmoins à la réalité³⁵. Plus précisément, il postule une « relation de similitude » sur des « mondes possibles », et propose qu'un contrefactuel a affaire aux « mondes »

³⁴Lewis (1978, p268 *sq.*)

³⁵Lewis parle plutôt d'actualité que de réalité, mais les enjeux de cette différence terminologique, étant d'ordre métaphysique, n'ont aucune importance pour le propos de ce travail.

où son antécédent est vrai qui sont les plus « proches », selon cette relation de similitude, du monde réel, ou comme il préfère de dire, du monde « actuel ». En l'occurrence, le contrefactuel est vrai si son conséquent est vrai dans tous ces mondes. En supposant certaines propriétés de cette relation, la théorie de Lewis implique une « logique » des contrefactuels, c'est-à-dire un certain nombre de « règles logiques » censées s'appliquer à eux³⁶. En motif du rapport entre fiction et contrefactuels, Lewis soutient qu'un texte de fiction a lui aussi affaire à un ensemble de mondes possibles qui sont distincts du monde réel mais néanmoins lui ressemblent, à savoir, *grosso modo*, les mondes possibles les plus proches du monde réel où les propos du texte sont vrais. Les vérités selon le texte de fiction sont, d'après cette approche, les énoncés qui sont vrais dans tous ces mondes les plus proches – tous ces « mondes du texte », pour ainsi dire.

Malgré son importance et son influence, l'analyse de Lewis n'est *qu'un exemple* de la mise en œuvre du rapport entre fiction et contrefactuels. Il y a, comme on vient de voir, un *parallélisme général et profond* entre fiction et contrefactuels, tenant aux propriétés communes de leurs rapports respectifs à la réalité, et donc une problématique commune dont s'occupe, dans un certain sens, la totalité de cette thèse. Le problème, dans les deux cas, est de *comprendre, dans les détails, la tension entre la distinction et la similitude qu'ils affichent par rapport à la réalité*. On se sert ici de ce parallélisme dans sa forme générale seule, sans supposition supplémentaire, et certainement sans vouloir soutenir aucune réduction de la fiction ou de la « vérité fictionnelle » aux contrefactuels, ou aux propriétés des mondes possibles. Au contraire, le parallélisme général entre fiction et contrefactuels dévoile une faiblesse dans l'analyse lewisienne des contrefactuels en termes de mondes possibles, dans la mesure où il permet de voir dans la théorie de Lewis l'équivalent, sur le plan de contrefactuels, des théories de la fiction qui adoptent la stratégie que l'on a qualifié de « mimétique » (§1.2.1). De même que la théorie du *faire comme si* et les autres théories discutées dans §§1.4 et 1.5, la théorie de Lewis pose *d'abord* une *distinction* tranchante avec la réalité, en l'occurrence la distinction entre le monde réel et les mondes « irréels » pertinents pour la fiction ou le contrefactuel, pour tenter *ensuite* de retrouver le rapport avec la réalité, en l'occurrence avec la notion de *similitude* entre mondes. Il s'ensuit que les critiques dressées contre l'application de la stratégie « mimétique » pour résoudre le problème de la fiction (§§1.4 et 1.5) s'appliquent également à l'égard de l'analyse des contrefactuels en termes de mondes possibles.

En particulier, la critique 3. de §1.4.1 a souligné le décalage entre la relative

³⁶Lewis (1973b), notamment Ch. 1, §§2.3 – 2.4 et Ch. 6.

stabilité de l'opposition entre *faire comme si* et croyance et la muabilité du concept de *faire comme si* et de son rapport à la croyance qui apparaît clairement dans les comparaisons entre différentes lectures d'un même texte par exemple. Cette critique s'applique également à la théorie lewisienne de contre-factuels et de la vérité fictionnelle. Cette théorie n'offre pas d'analyse des textes qui admettent des interprétations différentes, tel *Le Tour de l'écrou* de Henry James, où il y a, selon une lecture, des fantômes, mais non selon une autre lecture, qui voit dans les descriptions données des effets de l'imagination de l'un des personnages. Il n'est pas suffisant de soutenir que l'une des interprétations est vraie dans quelques mondes les plus proches, et l'autre est vraie dans d'autres mondes également proches³⁷, car il s'ensuit, par le fait que ni l'une ni l'autre des interprétations ne soit vraie dans *tous* les mondes les plus proches, que les deux sont fausses. Selon cette analyse, il ne serait ni vrai ni faux selon *Le Tour de l'écrou* qu'il y ait des fantômes derrière les événements racontés, dans la même manière qu'il n'est ni vrai ni faux que la gouvernante est de groupe sanguin A³⁸. Il faut donc reconnaître que les deux interprétations différentes ont affaire à deux *relations de similitude différentes*³⁹. Or, alors que Lewis admet des changements de la relation de similitude selon « la pratique et le contexte »⁴⁰, il n'offre que très peu de compréhension de ces changements et des rapports entre des relations de similitudes différentes. Dans le cas de la psychologie de la fiction, où les deux interprétations correspondent aux deux « modalités de sélection » des croyances ordinaires pertinentes dans la lecture du texte, le manque de compréhension des changements et des rapports entre ces différents « modes de sélection » comptait comme une faiblesse, dans la mesure où elles ont trait précisément à la relation problématique entre fiction et réalité, en l'occurrence entre *faire comme si* et croyance (§1.4.1). De manière semblable, il faut reprocher aux théoriciens des contre-factuels leur silence à l'égard des changements des relations de similitudes, dans la mesure où il implique un défaut dans leur compréhension du rapport entre contre-factuels et réalité.

³⁷Lewis permet qu'il y ait *plusieurs* mondes possibles également proches du monde actuel (Lewis, 1973b, §1.3), (Lewis, 1978, p269-270).

³⁸Lewis, *op cit.*, p270.

³⁹La différence entre les interprétations n'est pas prise en compte de manière satisfaisante si l'on identifie une interprétation (l'interprétation psychologique) comme vraie dans les mondes les plus proches du monde *actuel* et l'autre (l'interprétation supranaturelle) comme vraie dans les mondes les plus proches des mondes *que l'on s'est cru habiter il y a un siècle*, comme le fait Currie (1990, p66 *sq.*, mais voir sa note 24, p67). Car cela impliquerait que l'ambiguïté d'interprétation n'existe pas pour un lecteur actuel ni pour un lecteur contemporain à l'écriture du roman, alors qu'elle fait partie intégrante de l'œuvre.

⁴⁰Lewis (1981, p87 (§5)). Voir aussi 1973b, §4.2.

Cette critique, qui remonte effectivement à Goodman⁴¹, a des conséquences sur ce que l'on exige d'une théorie des contrefactuels. Il ne s'agit pas de trouver une sémantique particulière, telle une relation de similitude munie de certaines propriétés, et par conséquent une collection de règles logiques qui sont applicables à ou correctes pour les contrefactuels. Car les analyses qui se bornent à proposer une telle logique, à la façon de la théorie de Lewis, négligent les problèmes plus essentiels touchant aux changements des relations de similitudes, et plus généralement aux rapports entre les relations de similitudes à l'œuvre à des moments et à des contextes différents.

Le propos de ce travail n'est donc pas de discuter la plausibilité de telles ou telles propriétés des relations de similitude ou des sémantiques pour les contrefactuels qui ont été proposées dans la littérature, ni l'applicabilité ou la correction des règles logiques impliquées par ces propositions. Il consiste plutôt à développer une théorie à la mesure de modéliser les passages d'une sémantique, ou d'une relation de similitude, à l'autre.

Plus précisément, il faut distinguer entre deux genres de propriétés des contrefactuels dont il faut s'occuper. D'un côté, il y a des aspects « internes » qui valent dans des situations individuelles à des moments particuliers. En termes de l'analyse de Lewis, c'est l'ensemble de mondes possibles et la relation de similitude entre eux qui sont censés rendre fidèlement ces aspects. De l'autre côté, il y a des aspects « externes », qui ont trait aux rapports entre contrefactuels évalués dans des situations et à des moments différents. Tels sont des aspects de la dynamique de la relation de similitude qui sont souvent négligés par les théoriciens des contrefactuels.

Pour analyser les contrefactuels, il s'agira d'abord de définir un type de structure logique, qui sera *locale* et *vague* et qui permettra d'évaluer des contrefactuels dans une manière fidèle aux « propriétés internes » des contrefactuels, qui se montrent dans des situations individuelles. On appelle ces structures des *contextes*. En concevant les contextes vagues comme des ensembles ouverts au sens topologique, on arrivera à définir une topologie sur les contextes et donc une notion de *continuité* qui s'y appliquent. La théorie formulée en termes de ces contextes réussira à aborder la question essentielle de contrefactuels, touchant à leurs « aspects externes », dans la mesure où elle modélise les passages d'un contexte à l'autre de telle manière que ces passages apparaissent comme continus. Cette théorie incorporera donc aussi bien certains aspects traités par les théories traditionnelles que des aspects qu'elles délaissent, en les considérant être irréductiblement pragmatiques.

⁴¹Goodman (1954).

Cette théorie propose une *analyse* des contrefactuels, *non* une logique. Néanmoins, elle pourrait être comparée aux formalismes habituels consistant en des règles logiques censées appliquées aux contrefactuels, dans la mesure où l'on peut retrouver les règles logiques proposées dans la littérature, de trois manières différentes. Premièrement, les règles logiques peuvent correspondre à des propriétés structurelles des contextes ; deuxièmement, elles peuvent correspondre à des propriétés des transitions entre contextes ; et dernièrement, elles peuvent référer à des propriétés des structures globales qui « résument » les aspects appropriés de tous les différents contextes locaux. En mettant des contraintes appropriées sur les propriétés des contextes, des opérations sur eux, et les exigences sur une structure globale qui vise à les « résumer » respectivement, on retrouve les formalismes habituels. En ce sens, on peut dire que la théorie propose un *schème de logiques*, comportant certaines « variables d'ajustement », de telle sorte que, en les fixant de manière appropriée, on obtient les logiques de la tradition.

Enfin, il est clair que non seulement la problématique des contrefactuels ressemble à celle de la fiction, mais que la réponse apportée par cette théorie répète à maints égards les aspects de la réponse développée dans la première partie de cette thèse. La notion de contexte qui intervient dans les pages qui suivent est évidemment la contrepartie au plan logique de la structure instantanée de la conscience qui a été cruciale précédemment. L'approche de la dynamique, qui définit une notion de continuité à partir du vague des contextes pour montrer que les passages d'un contexte à l'autre sont continus, imite l'approche de la dynamique des structures instantanées de la conscience, qui s'est appuyée également sur leur caractère vague. Même la possibilité de retrouver les logiques traditionnelles en tant que propriétés des transitions entre contextes ou d'une structure qui « résume » en quelque sens tous les contextes rappelle celle de retrouver les États Mentaux des théories traditionnelles de la psychologie de la fiction en tant que propriétés des suites des structures de la conscience. L'approche des contrefactuels proposée prochainement, de même que l'approche de la lecture de la fiction adoptée précédemment, emprunte la stratégie que l'on a qualifiée de « naturaliste » (§1.2.3).

Dans le premier chapitre de cette partie (Ch. 4), le problème des contrefactuels sera introduit et approfondi, pour aboutir à un cahier de charges pour une éventuelle théorie des contrefactuels, comportant des aspects « internes » (§4.2) aussi bien que des aspects « externes » (§4.3). Au fur et à mesure que l'on repère des défis pour une théorie des contrefactuels, on constatera dans quelle

mesure les théories proposées réussissent à les relever. Dans le deuxième chapitre (Ch. 5), on développera une théorie des contrefactuels (§5.1) et évaluera son adéquation relative au cahier des charges établi précédemment (§5.2). Il sera enfin question des manières dont le cadre proposé par cette théorie permettra de retrouver les logiques habituelles des contrefactuels (§5.3 et l'Annexe A).

Chapitre 4

Déterrer le problème des contrefactuels

La deuxième partie de la thèse propose une théorie des contrefactuels. Cette théorie sera proposée dans le chapitre suivant ; le chapitre présent est préparatoire, visant à délimiter les objectifs de cette théorie. L’objet de la théorie est le contrefactuel au sens le plus vaste et le moins spécifié. À un tel niveau de généralité, le *problème* des contrefactuels est apparemment à portée de la main, mais effectivement plus difficile à préciser. Trop souvent, on s’élance dans une analyse sans le poser proprement. Ce chapitre a pour but de *bien cerner le problème des contrefactuels dans toute sa complexité*. Dit approximativement, ce problème est de comprendre le rapport entre le contrefactuel et la réalité. La première section introduit la notion de contrefactuel, son rapport à la fiction, et esquisse les grandes lignes de la question qui se pose à son égard. Dans les deux longues sections suivantes, cette question est approfondie, en repérant un certain nombre de phénomènes banals et intuitifs relatifs aux contrefactuels. On aboutira enfin à une liste de *desiderata* pour une théorie des contrefactuels. Une telle démarche lente et mesurée pourrait surprendre par son absence de maintes considérations des contrefactuels qui se trouvent dans la littérature ; pourtant, cette absence même la nécessite. Elle permettra de fixer le champ du questionnement actuel, pour le séparer d’autres questionnements moins réfléchis, et ainsi de préparer le développement d’une théorie dans la conscience des enjeux importants et de ceux qui ne le sont pas.

4.1 Premier tour : rapport à la fiction

Le but de cette section est de fixer l'objet de l'enquête qui suit, et d'esquisser la question qui sera posée à son égard. L'objet est le contrefactuel dans son sens le plus vaste et le moins spécifié. Le problème est de comprendre son rapport à la réalité. Il s'avère que ce rapport témoigne d'un parallélisme frappant avec le rapport entre la fiction et la réalité qui fut le sujet de la première partie de cette thèse. Ce parallélisme permettra l'emploi des constats, des techniques et des résultats antérieurs, à titre d'inspiration ou de suggestion, dans la considération ultérieure de contrefactuels.

4.1.1 Contrefactuels

Où la notion de contrefactuel, dans la généralité où elle sera traitée, est introduite.

Les contrefactuels ont suscité de vifs débats dans la philosophie analytique depuis plus d'un demi-siècle. Cet intérêt est sans doute lié à leur importance pour un grand nombre de questions épineuses relatives aux lois de la science, aux révisions et à la structure des croyances, au langage naturel, à la métaphysique et à la possibilité, à la fiction et la vérité fictionnelle. Il est bien possible de poser des questions séparées à l'égard des contrefactuels dans leurs domaines différents ; il est également possible d'entrer dans l'analyse des contrefactuels sans même spécifier quel genre de domaine, et donc de contrefactuel, on vise. L'intention de ce travail est *généraliste* : analyser le concept de contrefactuel dans sa plus grande généralité, en tant qu'il est employé et débattu dans toutes les perspectives mentionnées, et éventuellement d'autres. Aucun champ ou fonction spécifique de contrefactuels n'est supposé, et aucun recours à des champs ou fonctions particuliers ne sera permis dans la théorie de contrefactuels qui sera proposée. Il s'agit ici d'introduire le contrefactuel dans ce sens large, aussi bien que d'écarter certaines spécifications de la notion qui ne seront pas adoptées.

De manière générale, on s'accorde à identifier comme parangon de contrefactuel un conditionnel de la forme « si quelque chose était le cas, alors quelque chose d'autre serait le cas » ; par exemple « si cette allumette avait été grattée, alors elle se serait enflammée ». On désigne le contrefactuel avec antécédent A et conséquent B par « $A \Rightarrow B$ ». Jusqu'où étendre la notion de contrefactuel à partir de ce cas paradigmatique est une question débattue, et largement non résolue.

On privilégie parfois la distinction grammaticale, débouchant ainsi sur une notion de contrefactuel qui est intimement liée à son expression nécessaire dans

des *conditionnels subjunctifs* (aussi appelés irréels), et qui s'oppose à d'autres conditionnels, notamment des *conditionnels indicatifs*¹. Du reste, il y a une distinction envisageable selon la vérité de l'antécédent, selon laquelle un contrefactuel est un conditionnel dont l'antécédent est faux. On peut aller plus loin, en exigeant de plus que le conséquent le soit aussi. Ce faisant, on écarterait les dits « semifactuels »², qui s'apparentent à des contrefactuels à ceci près que leurs conséquents sont vrais : par exemple, « si l'allumette avait été grattée, elle ne s'allumerait pas » (énoncé au sujet d'une allumette mouillée) est un semifactuel.

Dans la discussion suivante, *aucune de ces distinctions ne sera posée au début*. Ce qui est entendu par le terme de contrefactuel se précisera au fur et à mesure que la question posée à leur égard est mise au point. En fin de compte, ce ne seront pas des propriétés linguistiques ou logiques qui importeront pour la notion de contrefactuel, mais plutôt le genre de *connexion* qu'il instaure ou affirme entre l'antécédent et le conséquent, et le rapport entre cette connexion et le monde. C'est, pour ainsi dire, par ses propriétés *effectives* qu'on identifie un contrefactuel.

4.1.2 Fiction et contrefactuels

Où le parallélisme entre la problématique de la fiction et celle des contrefactuels, tenant aux propriétés semblables de leurs rapports respectifs à la réalité, est mis en évidence.

Le problème de la fiction, qu'il soit posé en termes de ce qu'un texte de fiction autorise comme « vrai » – sa dite « vérité fictionnelle » – ou en termes des croyances d'un lecteur à son égard, est celui de son rapport au réel (éventuellement, ce qui est cru être réel). Qu'il s'agisse de la supposition de mortalité qui autorise la croyance qu'un personnage est en danger ou de la mise

¹Une division de la structure logique des conditionnels selon ces lignes grammaticales est soutenue notamment par Adams (1970) et Lewis (1973b, §1.1). Pour des opinions opposées voir par exemple Stalnaker (1975) ou Lowe (1995) qui soutiennent une parenté d'analyse mais non pas de « paramètres » figurant dans l'analyse.

²Le terme est introduit par Goodman (1954, p5), où il en reconnaît deux aspects ou « portées », dont tantôt l'un tantôt l'autre est souligné par les auteurs ultérieurs employant ce terme. D'une part, il y a la dite « portée pratique », accentuée par Goodman lui-même (p6) et plus évidente dans la forme « Même si... , toujours... » (« Même si l'allumette était grattée, elle ne s'allumerait toujours pas »), selon laquelle un semifactuel affirme (sans nécessairement signifier) $\neg(A \Rightarrow B)$ où B est vrai. D'autre part, il y a la « portée littérale » (note 2, p6), qui sera employée dans cet essai, selon laquelle un semifactuel est un contrefactuel $A \Rightarrow B$ où B est vrai. La formulation dans le texte, sous forme de « si..., alors... » plutôt que de « même si..., toujours... », est choisie puisqu'elle rappelle plutôt la parenté avec les autres contrefactuels qu'une différence d'intention illocutoire.

en jeu de connaissances topographiques pour autoriser des conclusions sur la position relative d'un personnage³, on semble s'appuyer sur des faits réels, hors texte, pour le comprendre, et pour en tirer ses « vérités fictionnelles ». À l'inverse, non seulement on peut parler de vérités fictionnelles hors du cadre de la lecture du texte, c'est-à-dire dans la réalité, mais on soutient également que la lecture de la fiction, et sans doute les vérités fictionnelles qu'un lecteur découvre dans sa lecture, influent sur (sa conception de) la réalité.

Ce même rapport difficile et complexe au réel est apparemment reproduit dans le cas de contrefactuels. D'une part, la vérité des contrefactuels dépend crucialement d'« autres » faits réels : si « si cette allumette avait été grattée, alors elle se serait enflammée » est vrai, c'est sans doute en raison de certains faits réels, dont le fait que l'allumette est sèche, qu'elle est bien faite, et qu'il y a assez d'oxygène. D'autre part, les contrefactuels, qui parlent apparemment des situations qui ne se produisent pas effectivement, jouent des rôles importants dans le monde réel et effectif : dans la science par exemple, ou dans la gestion des croyances⁴.

La problématique commune à la fiction et aux contrefactuels est celle de leur rapport à la réalité : celle de la contribution des faits réels à la fiction ou à la vérité (ou l'assertabilité) des contrefactuels, et inversement, de la pertinence de la fiction et des contrefactuels pour la réalité. Cette communauté est au cœur du parallélisme important entre les deux, lequel sera utilisé à plusieurs reprises, non seulement lors de la discussion de la question de contrefactuels, mais aussi bien dans le développement d'une théorie des contrefactuels. Dans un premier temps, on s'appuie sur la problématique de la fiction et l'analogie avec celle des contrefactuels pour préciser cette dernière. Dans un deuxième temps, on s'inspire de l'approche proposée pour le problème de la fiction en tant qu'il se pose dans le cas de sa lecture, et sur le rapport avec le problème des contrefactuels, pour formuler une approche au problème général des contrefactuels.

(Cette partie du travail, et notamment le chapitre suivant, contiendra des considérations générales relatives aux contrefactuels. On ajoute parfois, comme dans le cas présent, des détails particuliers ou techniques, pour ceux qui sont plus intéressés. Ces détails apparaîtront aux fins des paragraphes pertinents, dans une typographie plus petite. Un lecteur qui ne s'intéresse qu'aux grandes lignes de l'argument peut les sauter sans perte ; un lecteur qui désire des précisions supplémentaires aux propos du

³§1.1.1 ; Lewis (1978, p268).

⁴Voir, d'un côté Chisholm (1946) par exemple. De l'autre côté, le *locus classicus* est Ramsey (2003, p246) et par la suite Stalnaker (1968, p102), Harper (1975, p245) ; pour les complications, voir Gärdenfors (1986).

texte est susceptible de les trouver dans de telles postfaces techniques.)

C'est l'*analogie générale* entre contrefactuels et fiction, et notamment entre les structures de leurs rapports respectifs à la réalité, qui importera dans la discussion suivante. Pourtant, il convient de donner un exemple plus concret du rapport entre fiction et contrefactuel. Voici donc, de manière plus précise, la relation entre la question « psychologique » de la fiction, c'est-à-dire celle des *croyances* d'un lecteur à son égard, dont s'est occupée la première partie de ce travail, et la question « sémantique » des contrefactuels, ayant trait à leur *vérité*. On suppose, comme on le fait souvent dans la littérature, qu'il y a un opérateur intensionnel « dans la fiction f , ... » tel que les « vérités fictionnelles » d'un texte de fiction f sont ces E tels que « dans la fiction f , E ». L'intuition qui est répandue dans la littérature, du moins depuis l'article de Lewis⁵, consiste à analyser cet opérateur comme un contrefactuel. Or, il n'est pas correct d'analyser « dans la fiction f , E » directement comme le contrefactuel « si les énoncés de f étaient vrais, alors E serait vrai », en raison d'effets tels que l'insinuation, l'ironie, la non fiabilité du narrateur et la « méta-lecture » (où l'on tire des conclusions dont le narrateur n'est pas conscient à partir de ces paroles). Afin de contourner ce problème, on propose normalement de faire entrer un aspect humain ayant trait à la communication et à l'interprétation du texte, sous forme de la mise en jeu explicite du narrateur ou du lecteur, et le plus souvent de leurs croyances⁶. Nul besoin d'entrer dans les débats épineux sur le rôle et la nature du narrateur et du lecteur ; il suffit de se contenter avec une analyse de la forme suivante :

(**VF**) « Dans la fiction f , E » est vrai si et seulement si si un lecteur ordinaire recevait f dans des circonstances qui lui donnent à croire qu'elle était un fait connu, alors il serait raisonnable pour lui de croire E .⁷

Pourtant, la sensibilité de la vérité fictionnelle à l'égard des questions interprétatives semble attester une parenté spéciale entre la vérité fictionnelle et les croyances du lecteur à l'égard du texte. Plus précisément, il semble y avoir un *rapport étroit*, sinon une égalité stricte entre la formule ci-dessus, « logique » dans le sens où le contrefactuel prend portée large sur les croyances, et la formule suivante, « psychologique » dans la mesure où la croyance (ordinaire) prend portée large sur la vérité fictionnelle :

« Dans la fiction f , E » est vrai si et seulement s'il serait raisonnable

⁵Lewis (1978).

⁶Lewis (1978) fait entrer la notion de narrateur, Currie (1990) emploie une notion de narrateur et les croyances d'un lecteur (ordinaire), alors que Krasner (2002), pour sa part, emploie seulement les croyances du lecteur, comme dans l'analyse (**VF**) ci-dessous.

⁷Cette analyse est défendue dans Hill, *Quelques caractéristiques de la fiction*, §VIII (Mémoire de DEA d'Histoire et Philosophie des Sciences, Université de Paris 1, 2003).

*pour un lecteur ordinaire de croire que E est vrai selon f .*⁸

La formule **(VF)** représente d'une manière particulièrement aiguë le rapport entre les contrefactuels et la fiction, notamment la lecture de la fiction. Son importance dans ce travail est limitée à son titre d'exemple : à certains moments dans la discussion des contrefactuels, on se sert des exemples de vérités fictionnelles, comprises comme des contrefactuels. Néanmoins aucune *analyse* de la vérité fictionnelle ou de la lecture de la fiction en termes des contrefactuels n'est proposée ici, ni aucune réduction de l'une à l'autre ; c'est le rapport entre fiction et contrefactuels et lui seul qui importe.

4.1.3 La question des contrefactuels

Où l'on énonce le problème des contrefactuels comme celui de comprendre son rapport à la réalité. Le développement d'une théorie correcte des contrefactuels, apte à aborder ce problème, gagnera à l'établissement d'une liste de desiderata, laquelle puisera à diverses sources.

En deux mots, le problème de la fiction est celui de son rapport à la réalité. Or, comme on vient de le voir, le rapport entre le contrefactuel et la réalité affiche les difficultés pareilles : d'un côté, des faits de la réalité apportent des soutiens nécessaires aux contrefactuels, et de l'autre côté, des contrefactuels jouent un rôle important auprès des questions ayant trait apparemment à la réalité seule. La question que l'on pose ici à l'égard de contrefactuels est celle de *leur rapport à la réalité*.

Tout comme la notion de contrefactuel qui est en jeu ici, la question posée à son égard est *générale*. De même que dans l'enquête menée relativement à la fiction dans la première partie de ce travail, le manque de spécificité dans l'objet interdit l'emploi des notions appartenant à des domaines particuliers. Aucun sens ou fonction spécifique de contrefactuels, qu'ils soient métaphysiques, épistémologiques, logiques, linguistiques ou autres, n'est supposé ni privilégié ; aucun concept ressortissant à une fonction ou sens particulier de contrefactuel n'est admis dans la théorie des contrefactuels cherchée ici. De même que dans le

⁸Plus formellement, l'analyse **(VF)** est « $Rais(C_{LO}(f)) \Rightarrow Rais(C_{LO}(E))$ » ; le renversement de la portée de l'opérateur de croyance et du contrefactuel aboutit à « $Rais(C_{LO}(f \Rightarrow E))$ » (il est raisonnable pour un lecteur ordinaire de croire que, s'il recevait f comme un fait connu, E serait vrai) ; en rapprochant le fait de croire que E serait vrai si le texte était reçu comme un fait connu au fait de croire E selon f , on obtient « $Rais(C_{LO}(E \text{ selon } f))$ », l'analyse formulée dans le texte.

cas de la fiction, s'il fallait qualifier la question posée, elle serait *méthodologique* : elle appelle comme réponse une *manière de concevoir les contrefactuels qui est propice pour les penser dans leur rapport à la réalité*.

Notamment, le problème posé ici consiste à *comprendre* les contrefactuels et leur rapport à la réalité, non pas à *imposer* des contraintes sur des contrefactuels et ce rapport. En ce sens, la problématique est de prime abord *descriptive* plutôt que *normative*. Par conséquent, on ne saurait pas supposer à cette étape préliminaire de l'enquête que la théorie des contrefactuels cherchée prenne la forme d'une logique au sens traditionnel, c'est-à-dire qu'elle consiste à proposer un ensemble de *règles*. Une *analyse* des contrefactuels n'est pas nécessairement une *logique* des contrefactuels. Car, si l'on considère que les règles logiques *imposent* des *contraintes* sur les contrefactuels, permettant certains comportements logiques et interdisant d'autres, c'est-à-dire si l'on considère qu'elles soient *normatives*, alors on ne prétend pas qu'elles apportent une réponse à la question descriptive posée ici. Si, en revanche, on considère que les règles logiques décrivent le comportement des contrefactuels, alors il s'agit d'une thèse qui sera confirmée ou réfutée par une étude approfondie à leur égard, et corrélativement admise ou rejetée par la théorie qui résulte de cette étude, et non une précondition de cette théorie même. Évidemment, si elle renonce aux règles logiques proposées dans la tradition, une théorie complète des contrefactuels devrait néanmoins expliquer le rapport entre ces règles logiques et les faits relatifs aux contrefactuels qu'elle prétend saisir de manière plus exacte. Elle devrait expliquer les règles logiques proposées, en tant que « phénomènes » concernant des contrefactuels, si l'on peut dire.

Avant de développer une théorie qui pourrait prétendre à une compréhension des contrefactuels, compris dans un tel sens large, il faut expliciter plus précisément les *desiderata* pour une telle théorie. Pour établir une telle liste de *désiderata*, il s'agit d'adapter la méthodologie simple selon laquelle on ramasse autant de « données » que possible. Tout préjugé sur la forme de ces données ou de leur source sera banni : l'intuition quotidienne à l'égard des contrefactuels aussi bien que des acquis des théories sophistiquées seront également acceptés, non seulement les propriétés « logiques » mais aussi les phénomènes qui ne se donnent pas immédiatement à l'expression rigoureuse joueront un rôle. Une illustration de l'étendue de l'ensemble des données dont une théorie des contrefactuels devrait rendre compte vient des questions relatives à certains contrefactuels qui ne sont pas tranchées dans l'intuition. Le contrefactuel « si cette allumette avait été grattée, Jacques Chirac serait le président de France en 2004 » est-il vrai ou non ? On est embarrassé ici, trébuchant entre deux intui-

tions opposées. *Cette hésitation même* sera considérée comme un fait à l'égard des contrefactuels qui est à comprendre. En le traitant ainsi, on s'oppose à la plupart des théoriciens de la tradition, qui non seulement font écrouler cette subtilité, en se prononçant en faveur d'un côté ou de l'autre, mais le font largement en fonction des besoins de leurs propres théories. Ce qui relève des pires des corruptions méthodologiques : le fait que l'intuition ne tranche pas et ne semble pas pouvoir trancher ne leur permet pas de décider à sa place, il ne leur permet pas de décider les faits pour cadrer avec leurs théories.

Le travail de considérer soigneusement les défis pour une théorie de contrefactuels n'a rien de trivial. Au contraire, ce travail, cette précision de la *tâche* d'une théorie des contrefactuels, est trop souvent négligé, avec des conséquences fâcheuses, de telle sorte que la liste de *desiderata* à laquelle on aboutira aura une valeur en elle-même. Cette liste esquissera le champ conceptuel dans lequel la théorie des contrefactuels sera développée. L'argumentation soutenant cette liste peut donc être également conçue comme un plaidoyer pour ce champ conceptuel, qui le compare d'ailleurs à d'autres champs conceptuels et aux théories qu'ils abritent.

Les deux sections suivantes sont consacrées à ce travail important, c'est-à-dire aux phénomènes détaillés ayant trait aux contrefactuels ; elles permettront d'esquisser un questionnement, un champ, une tâche pour une éventuelle théorie des contrefactuels.

4.2 Second tour : intuitions, données et défis

La question générale des contrefactuels est la question de leur rapport à la réalité. Dans cette section et la suivante, on raffine et approfondit cette question en établissant une liste précise de *desiderata* pour une théorie des contrefactuels. On la puise de deux sources différentes : d'abord, on se tourne vers les théories proposées des contrefactuels, qui, par les aspects qui leur sont communs, indiquent quelques propriétés évidentes et communément admises des contrefactuels. Ensuite, on se met à l'écoute de l'intuition quotidienne à l'égard des contrefactuels, en repérant les avis primitifs et immédiats à leur égard. Bien entendu, ces intuitions « crues » ne sont que rarement rigoureuses et précises, de telle sorte que la discussion suivante, qui, dans un premier temps, se contente de repérer ces intuitions sans trop les analyser, sera largement non formelle et non rigoureuse, figurant des termes intuitivement compris mais non pas précisément définis.

La constitution de cette liste d'aspects « de base » de contrefactuels aura

deux fins supplémentaires. La première est critique : en remarquant certaines propriétés intuitives des contrefactuels, et la manière dont les théories existantes en rendent compte, on met ces théories à l'épreuve. La seconde est constructive : les aspects et les thèmes soulignés ici serviront à préparer le terrain pour l'élaboration d'une théorie des contrefactuels dans le chapitre suivant. Cette théorie rendra compte de la plupart de ces aspects, en les accordant des sens rigoureux et précis ; à l'inverse certains de ces aspects joueront des rôles importants dans l'élaboration et la motivation de cette théorie. En bref, le chapitre actuel et le chapitre suivant se complètent : le premier pose les défis, le dernier les relève. Il ne convient pas d'énumérer actuellement ces défis, ni les manières d'en rendre finalement compte ; ils sont résumés respectivement dans §§4.3.4 et 5.2.4.

4.2.1 Intuitions traditionnelles

Où sont présentés les faits de base relatifs aux contrefactuels, admis par toute théorie à leur égard. De manière générale, un contrefactuel peut être pensé comme une conséquence stricte figurant des prémisses sous-entendues : son conséquent est conséquence de son antécédent, pris ensemble avec ces prémisses. Comme tel, un contrefactuel comporte une part « structurelle » – la conséquence stricte – et une part « matérielle » – les prémisses sous-entendues. Le tri exact entre ces parts est sujet aux débats ; chacun est cependant d'accord que la notion de conséquence et les prémisses qui sont pertinentes pour un contrefactuel particulier proviennent de la pragmatique.

Selon une conception répandue, les contrefactuels sont « à mi-chemin » entre deux « sortes » d'implication bien connues, à savoir *l'implication matérielle* (parfois appelée implication philonienne), $A \supset B$, qui est vraie si et seulement si A est faux ou B est vrai, et *l'implication stricte*, $A \rightarrow B$, qui est vraie si et seulement si B est une « conséquence » de A . On s'accorde également à ce que le point exact occupé par les contrefactuels entre ces deux implications est, pour ainsi dire, fixé par un *apport pragmatique*. L'exposition qui suit de cette conception répandue des contrefactuels procède en trois temps. On explore d'abord les notions d'implication stricte et d'implication matérielle et deux manières différentes de les concevoir. Ensuite, on expose le genre d'analyse souvent proposée du contrefactuel et les deux manières de concevoir le contrefactuel qui découlent des différentes conceptions d'implications matérielles et strictes. On vient enfin aux apports pragmatiques communément supposés dans les analyses des contrefactuels.

Implication matérielle et implication stricte

Dans la littérature, on entame souvent une analyse des contrefactuels en les comparant à deux autres notions bien connues d'implication, à savoir l'*implication matérielle* et l'*implication stricte*. L'implication matérielle $A \supset B$ est vraie si et seulement si A est faux ou B est vrai. L'implication stricte $A \rightarrow B$ est vraie si et seulement si B est *conséquence* de A , ou, comme on dit souvent, si la proposition exprimant l'implication matérielle, « A est faux ou B est vrai », est vraie *nécessairement*. Ces deux manières de définir l'implication stricte – en termes de conséquence et en termes de nécessité – sont techniquement équivalentes : pour une notion de nécessité donnée, on pourrait définir une notion de conséquence comme nécessité de l'implication matérielle ; à l'inverse, pour une notion de conséquence donnée, il y a la notion de nécessité correspondant, selon laquelle une proposition est nécessaire si elle est conséquence du vrai. Néanmoins, on arrive à des conceptions de l'implication stricte qui diffèrent, non pas techniquement mais *heuristiquement*, selon que l'on met en avant la nécessité ou la conséquence. Il en découle des conceptions différentes du rapport entre l'implication matérielle et l'implication stricte et, dans la mesure où un contrefactuel est pensé comme une sorte de « mélange » des deux implications, deux conceptions différentes des contrefactuels.

L'une des conceptions s'inspire d'une image « *extensionnelle* » de la nécessité : on se donne d'abord un ensemble de points, intuitivement des « mondes (possibles) », des « situations » ou des « états » ; on conçoit ensuite une proposition comme un ensemble de tels points – intuitivement l'ensemble de « mondes », « situations » ou « états » où la proposition est vraie ; il en découle qu'une proposition est nécessaire si et seulement si elle est vraie dans tous les mondes (situations, *etc.*). Puisqu'une proposition est un ensemble de mondes, il n'y a qu'une proposition nécessaire : l'ensemble de tous les mondes. Une implication stricte vraie est une implication matérielle nécessaire, c'est-à-dire une implication matérielle qui est vraie dans tous les mondes. De l'autre côté, on désigne un des mondes comme le monde « *actuel* » : une implication matérielle est vraie si elle est vraie dans ce monde.

Selon l'autre conception – que l'on peut qualifier d'« *algébrique* » – au lieu de se concentrer sur les éléments « de base », on s'intéresse aux propositions elles-mêmes et aux rapports entre elles. Dans ce cas, on conçoit une structure reliant ces éléments – les propositions – avec les opérations sur eux, telle la conjonction de propositions, et notamment un ordre sur eux (\leq), lequel représente la conséquence. Cette structure est une *algèbre de Boole* (pour la logique proposi-

tionnelle classique)⁹. Dans une telle structure, pour deux propositions A et B , $A \leq B$ signifie que B est conséquence (stricte) de A ; l'implication stricte $A \rightarrow B$ est vrai si $A \leq B$. Il y a un élément au-dessus de (c'est-à-dire conséquent de) tous les autres – \top . Une proposition est nécessaire dans cette structure si elle est une conséquence du \top ; or, par définition, \top est une conséquence d'elle, donc toute proposition nécessaire est, pour ce qui s'agit de cette structure, égale à \top . Une *valuation* sur une algèbre de Boole assigne à chaque élément une valeur de vérité – Vrai ou Faux – de manière consistante ; une implication matérielle $A \supset B$ est vraie (selon une valuation) si l'élément $\neg A \vee B$ ¹⁰ est accordé la valeur V.

Ce sont deux conceptions différentes, mais les *structures* qu'elles instaurent sur les propositions sont équivalentes, c'est-à-dire, pour parler techniquement, isomorphes. De fait, on sait bien traduire entre eux : grossièrement, une proposition comme élément algébrique est l'équivalent de la proposition comme ensemble de mondes, et un monde est la conjonction de toutes les propositions (comme éléments) qui y sont vraies (dont il est membre de l'ensemble correspondant)¹¹. La différence entre les deux présentations n'est donc pas *structurelle* importante. Si l'on travaille avec des mondes, on peut construire l'algèbre de Boole de ses propositions. D'autre part, si l'on travaille avec une algèbre de Boole, on pourrait en tirer un ensemble de mondes. Soulignons cependant que si dans l'algèbre il n'y a qu'un nombre limité de propositions, plutôt que *toutes* les propositions d'un langage donné, alors il y aura des propositions de ce langage qui *n'ont pas de valeur de vérité* dans *aucun* des mondes de la conception « extensionnelle » tirée de cette algèbre. Ils ne sont des mondes possibles dans aucun sens métaphysiquement propre (s'il y en a un), mais plutôt des « petits mondes possibles »¹². Puisque la question métaphysique de mondes possibles n'est pas important ici, on continuera de les appeler mondes.

La comparaison entre les deux conceptions est résumée dans le tableau 4.1.

Au-delà de l'isomorphisme structurel, les intuitions offertes par les deux conceptions ainsi que les approches recommandées des questions logiques, diffèrent. La conception extensionnelle exige que pour toute manipulation on

⁹Pour la définition technique de l'algèbre de Boole, voir §5.1.1, Définition 5.1.1.

¹⁰Lu : non A ou B .

¹¹En vérité, les algèbres de Boole sont plus générales que les mondes possibles dans la mesure où il y a des algèbres de Boole sans atomes (c'est-à-dire, sans mondes possibles équivalents). Puisque l'on travaille ici généralement avec des algèbres de Boole finies et donc avec des atomes, on ignore cette complexité dans la suite. On trouve des détails relatifs à ces points dans n'importe quel bon manuel de logique.

¹²Terme apparemment introduit par Savage (1954).

TAB. 4.1 – Conceptions algébrique et extensionnelle

Conception algébrique	Conception extensionnelle
Algèbre de Boole	Ensemble d'ensembles de « mondes »
Valuation des éléments de l'algèbre	Désignation d'un monde « actuel »
Implication matérielle = élément de l'algèbre ($\neg A \vee B$)	Implication matérielle = ensemble de mondes
Implication matérielle vraie si l'élément $\neg A \vee B$ évalué comme vrai	Implication matérielle vraie si le monde actuel appartient à l'ensemble approprié
Implication stricte = propriété structurelle ($A \leq B$)	Implication stricte = implication matérielle vraie dans tous les mondes

passe par les mondes. À l'opposé, la conception algébrique permet de travailler directement avec les propositions elles-mêmes. Du coup, ces perspectives *conçoivent* la différence entre l'implication matérielle et l'implication stricte de manière légèrement différente. Dans la conception extensionnelle, la différence est celle entre *l'un* et *le tout* : une implication matérielle $A \supset B$ est vraie si *non* A ou B est vraie dans *le monde actuel* ; une implication stricte $A \rightarrow B$ est vraie si *non* A ou B est vraie dans *tous les mondes*. Dans la conception « algébrique » la différence est celle entre un *élément* et un *aspect de la structure* : une implication matérielle $A \supset B$ est *l'élément* $\neg A \vee B$, elle est vraie si l'on *assigne* à cet élément (proposition) la valeur vrai. Par contraste, une implication stricte $A \rightarrow B$ est vraie si $A \leq B$ dans l'algèbre : elle exprime un fait sur la *structure de l'algèbre*.

Les différentes intuitions offertes par ces différentes conceptions des implications matérielle et stricte entraîneront des perspectives différentes sur le contre-

factuel. Ils serviront à plusieurs reprises dans la suite d'outils pour exprimer et comprendre divers phénomènes relatifs aux contrefactuels.

Notion de conséquence et prémisses

Le chemin que l'on emprunte souvent pour motiver l'analyse de contrefactuels comporte trois étapes. D'abord, on remarque que les contrefactuels ne sont pas les implications matérielles. Car, alors que l'implication matérielle $A \supset B$ est vraie si A est faux, un contrefactuel avec un antécédent faux ne doit pas sa vérité qu'à la fausseté de l'antécédent. Le contrefactuel « si l'allumette avait été grattée, elle ne s'enflammerait pas », par exemple, est faux.

Ensuite, on constate que les contrefactuels ne sont pas non plus les implications strictes. Car le lien affirmé par un contrefactuel entre son antécédent et son conséquent n'est apparemment une nécessité dans aucun sens « métaphysique » ou fort du terme : il se pourrait bien que l'allumette ne s'enflamme pas après être grattée, alors que le contrefactuel correspondant était vrai.

Enfin, raisonne-t-on, si la vérité d'un contrefactuel ne consiste ni dans la vérité d'un simple énoncé, comme dans le cas de l'implication matérielle, ni dans l'existence d'une relation de conséquence, comme dans le cas de l'implication stricte, alors il doit consister en une relation de *conséquence* qui met en jeu, outre que l'antécédent explicite, *certain énoncés vrais* (selon la valuation ou la désignation du monde actuel). En appelant la *notion de conséquence* N , et les énoncés ajoutés à l'antécédent les *prémisses* P_i , on obtient, dans sa forme la plus générale, l'analyse **(AGC)** :

(AGC). $A \Rightarrow B$ est vrai si et seulement si B est conséquence de A pris ensemble avec des prémisses P_i , selon la notion de conséquence N .

Le contrefactuel est donc bien une implication « hybride » entre l'implication stricte et l'implication matérielle, dans la mesure où, comme l'implication stricte, elle fait entrer une notion de conséquence, et comme l'implication matérielle, elle s'appuie sur la simple vérité de certains énoncés. Par conséquent, le contrefactuel se comprend également dans la conception extensionnelle que dans la conception algébrique, quoique avec des différences de style et d'intuition selon la conception.

Du point de vue de la conception extensionnelle, il est question de l'ensemble de mondes où il faut que l'implication matérielle $A \supset B$ soit vraie pour que le contrefactuel $A \Rightarrow B$ soit vrai. Exiger $A \supset B$ dans le monde actuel est trop faible : c'est le cas de l'implication matérielle. Exiger $A \supset B$ dans tous les mondes est trop fort : c'est le cas de l'implication stricte. Il faut l'exiger dans un

certain ensemble de mondes, ensemble qui comporte plus que le monde actuel mais moins que tous les mondes¹³. Cet ensemble est une proposition dans le sens de cette conception ; elle est donc la prémisse dans l'analyse. Et la notion de conséquence est celle qui correspond à la notion de nécessité impliquée par l'ensemble de tous les mondes.

Par ailleurs, du point de vue de la conception algébrique la particularité des contrefactuels ressort plus clairement. Un contrefactuel n'est ni un élément de l'algèbre, comme l'implication matérielle, ni un aspect de sa structure, comme l'implication stricte. Il comporte *à la fois un aspect « matériel »*, référant à des éléments, *et un aspect « structurel »*, ayant trait aux rapports entre les éléments de l'algèbre. Les prémisses sont des énoncés vrais selon la valuation sur l'algèbre ; elles sont notamment des *éléments de l'algèbre*. La notion de conséquence correspond à la *structure de l'algèbre* ; le rapport entre la conjonction de l'antécédent et les prémisses avec le conséquent est un *aspect de cette structure*.

L'importance de cette forme général d'analyse est sans doute reliée au fait que, *a priori*, elle évite les dites « sophismes contrefactuels »¹⁴. Il y a un certain nombre de règles d'inférence qui sont valides pour l'implication matérielle ou l'implication stricte mais apparemment pas pour les contrefactuels. Par exemple, on est normalement d'accord sur le fait que la *contraposition* – la règle selon laquelle $A \Rightarrow B$ entraîne $\text{non } B \Rightarrow \text{non } A$ – n'est pas valide pour les contrefactuels : de « si l'allumette avait été grattée (A), alors elle s'enflammerait (B) », il ne s'ensuit pas que « si l'allumette ne s'enflammait pas ($\text{non } B$), alors elle n'aurait pas été grattée ($\text{non } A$) ». Elle n'est pas non plus valide, *a priori*, selon la forme générale (**AGC**). $A \Rightarrow B$ est vrai si B est conséquence, selon une notion N , de A pris ensemble avec des prémisses P_i ; or, de $\text{non } B$, il s'ensuit seulement que tantôt A tantôt une des prémisses P_i n'est pas vraie (en supposant la même notion de conséquence N), alors le dernier contrefactuel n'est pas nécessairement vrai. Plutôt que d'admettre que « si l'allumette ne s'enflammait pas ($\text{non } B$), alors elle n'aurait pas été grattée ($\text{non } A$) », on pourrait renvoyer le fait contrefactuel ne pas s'enflammer à un défaut dans la fabrication de l'allumette ($\text{non } P$). Ou, pour prendre un autre exemple, le *renforcement de l'antécédent*, selon laquelle $A \Rightarrow B$ implique $A \& C \Rightarrow B$, n'est pas valide pour les contrefactuels : « si l'allumette avait été grattée (A), alors elle s'enflammerait (B) » est admissible, tandis que « si l'allumette avait été grattée (A) et elle n'était pas sèche (C), alors elle s'enflammerait (B) » ne l'est pas. (**AGC**) est fidèle à cette propriété : puisque (supposant encore une notion de

¹³C'est à peu près la logique de Lewis (1973b, §§1.2–1.3).

¹⁴Pour une discussion approfondie, voir Lewis (1973b, §1.8).

conséquence N fixe) les prémisses pertinentes pour le premier contrefactuel ne sont pas nécessairement celles qui sont pertinentes pour le second, la vérité du premier contrefactuel n'implique pas celle du second. Par exemple, si *non C* figure parmi les prémisses P_i pour $A \Rightarrow B$ (comme dans l'exemple donné ci-dessus), il n'est pas parmi les prémisses pour $A \& C \Rightarrow B$ (sous peine de contradiction), et la vérité du dernier n'est pas impliqué par celle du premier. Tout évidemment, pour un choix particulier et astucieux de prémisses et de notion de conséquence, on pourrait faire valider ces règles, mais de manière générale, elles sont invalides selon **(AGC)**. Pour cette raison, on se permettra de faire l'économie de la considération de telles règles pour chacune des analyses individuelles qui réalise cette forme générale.

Du reste, la majorité écrasante des analyses disponibles, quoiqu'elles diffèrent sur la forme de la notion de conséquence et des prémisses, sont des exemples de la forme générale **(AGC)**. Quelques exemples suffisent pour illustrer la diversité des analyses proposées ainsi que leur communauté.

Un exemple explicite de ce genre d'analyse est celle proposée par Goodman¹⁵, qui distingue la prémisse et la conséquence – il les appelle les « conditions pertinentes » et les « principes de connexion » – et se pose la question de leur détermination.

D'un autre côté, le genre d'analyse la plus répandue se conçoit en termes des « mondes possibles ». Selon ce genre d'analyse, pour le dire grossièrement, on pose une *relation de similitude comparée* relative à chaque monde possible i , et on dit qu'un contrefactuel $A \Rightarrow B$ est vrai dans le monde i si et seulement s'il y a un monde où A et B sont vrais est vrai qui est plus proche de i (selon l'ordre de similitude à i) que tout monde où A est vrai et B est faux¹⁶. Or, comme le reconnaît Lewis¹⁷, cette analyse s'exprime également sous forme **(AGC)**, c'est-à-dire en termes de prémisses et d'une notion de conséquence. Pour évaluer un contrefactuel $A \Rightarrow B$ dans un monde i (le monde « actuel »), on prend comme prémisse la proposition qui est l'ensemble de mondes plus ou également proches du i que le(s) monde(s) le(s) plus proche(s) où A est vrai. La notion de conséquence est celle qui correspond à la notion de nécessité circonscrite

¹⁵Goodman (1954, Ch. 1).

¹⁶Lewis (1973c, p10). On dépouille l'analyse de Lewis de ses réserves et de ses propriétés non pertinentes pour alléger l'expression.

¹⁷Lewis (1973c, §2.2; 1973b, §2.6) (on simplifie son traitement dans des manières qui ne sont pas essentielles pour les propos de cette discussion). Cette manière de mettre l'analyse lewisienne dans la forme **(AGC)**, et notamment ce traitement des prémisses, n'est pas la seule; le débat autour les diverses possibilités n'est cependant pas important ici (voir par exemple Lewis (1973b, p70), Bennett (1974, §6)).

par tous les mondes possibles qui entrent dans l'ordre de similitude. Alors que l'on se servira de l'analyse en termes de similitudes dans la suite, les détails de la relation de similitude, lesquels sont débattus, ne seront pas de grande importance ici ; dans tous les cas, l'analyse pourrait être mise sous la forme (AGC).

Du reste, il y a des analyses tout à fait parallèles qui n'emploient pas la logique classique mais d'autres logiques, telles la logique pertinente (une logique avec une forme d'implication plus forte). Mares et Fuhrmann¹⁸ remplacent la relation de similitude comparée par une fonction qui prend toute paire d'une proposition et un monde à un ensemble de mondes, lequel contient les analogues des mondes les plus proches où la proposition est vraie. Le contrefactuel est valide si l'implication pertinente (plutôt que matérielle) est vraie dans ces mondes. L'ensemble de tous les mondes fournit la notion de nécessité comme ce qui est vrai dans tous les mondes, et ainsi la notion de conséquence. Les propositions vraies partout dans l'ensemble qui est l'image de la fonction sont des prémisses auxquelles se joignent à l'antécédent pour impliquer le conséquent selon la notion de conséquence¹⁹.

Il y a enfin des analyses qui laissent tomber la prémisse ou la relation de similitude, en transférant le poids entièrement sur la notion de conséquence. Le cas le plus explicite est la théorie de Lowe²⁰, selon laquelle le contrefactuel « si A était le cas, B serait le cas » est à analyser, à quelques détails non essentiels près, comme « il est 'nécessaire' que A est faux ou B est vrai » ; c'est-à-dire, vu le rapport entre nécessité et conséquence, comme « B est conséquence de A ».

On trouve également dans cette catégorie un courant probabiliste hésitant, qui suggère d'analyser l'*assertibilité* (*assertibility*) du contrefactuel $A \Rightarrow B$ comme la probabilité de B conditionnée sur A , $p(B/A)$ pour une fonction de probabilité *relative à un moment passé hypothétique approprié*²¹. Dans ce cas, l'aspect structurel consiste en les valeurs des probabilités conditionnelles (les rapports entre les propositions) ; l'aspect matériel consiste en les probabilités des propositions (la valeur que l'on accorde aux propositions individuelles). De

¹⁸Mares & Fuhrmann (1995). Mares (2004) donne une autre analyse d'inspiration pareille, mais avec d'importantes différences.

¹⁹Dans ce cas, il ne s'agit pas de prendre la conjonction des prémisses et de l'antécédent : on a plutôt $P \supset (A \rightarrow_p B)$ où P est la conjonction de l'ensemble de prémisses, et \rightarrow_p est l'implication pertinente.

²⁰Lowe (1995).

²¹Voir Adams (1975, Ch. IV) et Edgington (1995, §10). Le résultat de Lewis (1975) implique l'impossibilité d'une analyse de la *vérité* d'un contrefactuel de cette manière, en telle sorte que certains théoriciens, tels Adams et Edgington, ont échangé la question de la vérité des conditionnels (dont les contrefactuels) contre la question de leur *assertibilité*.

sorte que la probabilité conditionnée à l'œuvre correspond à la conséquence et il n'y a pas de prémisse, comme dans le cas de Lowe.

Part non formelle

L'implication matérielle est *formelle* – la vérité de $A \supset B$ est une fonction des valeurs de vérité de ses composants (A et B) ; en revanche, la vérité d'un contrefactuel dépend d'une certaine donnée supplémentaire aux énoncés composants (A et B). Comme le dit Lewis, la forme d'analyse (**AGC**) et les analyses qui en sont des exemples ne sont que des « squelettes » qui demandent à être étoffés, notamment par une notion de conséquence et des prémisses²². Ce qui veut dire, selon les cas, par des « conditions pertinentes » et des « principes de connexion », par une relation de similitude comparée, par une notion de nécessité ou par une fonction de probabilité. Dans la mesure où le contrefactuel n'est pas formel, il comporte une part *non formelle* ; dans la mesure où il faut la donnée d'une notion de conséquence et des prémisses (quelles que soient leurs formes), celles-ci constituent sa part *non formelle*. Non seulement toutes les théories des contrefactuels comportent une telle part non formelle, mais toutes les théories admettent que la part non formelle, c'est-à-dire la notion de conséquence et les prémisses employées dans l'analyse, pourrait se modifier selon le temps ou la situation. De plus, beaucoup d'entre elles sont d'accord pour voir dans cette part non formelle un élément ressortissant à la pragmatique.

Parmi les théoriciens mentionnés ci-dessus, Goodman est le seul à afficher une volonté d'aborder proprement la question de cette part non formelle : c'est le thème des derniers chapitres de son *Fact, Fiction and Forecast* (1954), où il traite cette question sous l'enseigne de la question des lois²³. Puisque son analyse de lois admet que les énoncés qui ont le statut des lois peuvent changer entre des instants éloignés dans le temps²⁴, la notion de conséquence et les prémisses impliquées dans l'analyse ne sont pas fixes, mais peuvent varier entre situations.

Désormais, cette variation est devenue un élément permanent des analyses des contrefactuels, où il vient être le plus souvent assimilée à la question du « contexte » – c'est-à-dire à une question relevant de la *pragmatique*. Comme le dit clairement Lewis, différentes relations de similitude comparée « sont appropriées à des contextes différents »²⁵. Stalnaker et Lowe sont explicites sur

²²Lewis (1979b, p465).

²³*Op cit.*, Ch.II sur le rapport entre contrefactuels et sa question de projectibilité, §III.4 sur le rapport avec la question des lois (*lawlikeness*) et Ch. IV sur l'approche suggérée.

²⁴C'est le sens de la dépendance « circulaire » de la validité d'une règle d'induction sur les inférences passées ; voir *op cit.*, Ch III, §2.

²⁵*Ibid.*

ce point ; l'analyse du dernier joue d'ailleurs sur la sensibilité de la notion de conséquence employée dans l'analyse au contexte²⁶. Quant aux auteurs qui sont moins explicites, tel Mares & Fuhrmann, Adams, et Edgington, chacun suppose son appareil théorique – relation de similitude ou fonction probabiliste – sans indication de sa provenance ou de ses propriétés non logiques. En bref, la part non formelle est reconnue comme un *apport pragmatique, relatif à, fourni par ou provenant du « contexte »*. Il conviendra d'explicitier la donnée du contexte dans le formalisme : on écrira donc $\vdash_C A \Rightarrow B$ pour « le contrefactuel « si A , alors B » est évalué comme vrai dans le contexte C » (c'est-à-dire, selon la notion de conséquence et les prémisses fournies par C).

Une remarque exégétique importante s'impose en conclusion : les parts non formelles supposées par les analyses considérées jusqu'à présent affichent généralement une certaine force, à savoir, elles sont *suffisantes à évaluer tout contrefactuel du langage*. Par exemple, une relation de similitude entre mondes est telle que pour tout A non contradictoire, il y aura un monde le plus proche du monde i où A est vrai, et donc on peut évaluer tout contrefactuel $A \Rightarrow B$ en regardant la valeur de B à ce monde et à d'autres mondes également proches²⁷. Pour ainsi dire, la donnée d'une relation de similitude fixe la notion de conséquence qui s'applique dans l'analyse de tous les contrefactuels et fournit les prémisses pour l'analyse de n'importe quel contrefactuel, en fonction de son antécédent. Ou, autre exemple, une fonction probabiliste accorde une valeur de probabilité à B conditionnée sur A pour tout A et B , et donc évalue tout contrefactuel $A \Rightarrow B$. Même Goodman, qui parle explicitement en termes de prémisses et de conséquences, entend donner une caractérisation des prémisses telle que, dans une situation donnée, la caractérisation fournit les prémisses nécessaires pour évaluer chaque contrefactuel²⁸.

Pour E un ensemble d'énoncés, et T une théorie qui évalue ces énoncés, on dira que T est *globale_E* si elle évalue tous les énoncés de E : à chaque énoncé elle assigne tantôt la valeur vrai tantôt la valeur faux (tantôt la valeur indéterminé ou n'importe quelle autre valeur qui est permise par la théorie). Pour chacune des théories de contrefactuels considérée ci-dessus, T , il y a un ensemble fixe de contrefactuels CL (tous les contrefactuels du langage), tel qu'il y a, pour chaque contexte particulier C , l'appareil (relation de similitude, fonction probabiliste, notion de conséquence et prémisses) permettant à T d'accorder à *tout* contre-

²⁶Lowe (1995, §VII).

²⁷On fait abstraction des questions relatives aux possibilités des mondes possibles qui n'entrent pas dans la relation de similitude et aux contrefactuels à antécédent contradictoire, qui sont toujours évaluées (comme vrai), Lewis (1973b, p16), Stalnaker (1968, p170).

²⁸*Op cit.*, Ch I, §2.

factuel de CL une valeur de vérité dans \mathcal{C} . Pour chaque \mathcal{C} , T est *globale* _{E} . Dans ce sens, on pourrait qualifier ces analyses des contrefactuels de *globales*.

4.2.2 La frontière imprécise et arbitraire

Où l'on constate que la frontière qui divise les prémisses de la notion de conséquences est imprécise et arbitraire.

Étant donné la forme générale (**AGC**), entendue comme un rassemblement de quelques intuitions fondamentales à l'égard des contrefactuels, il reste à l'approfondir et à la développer tant que possible. C'est-à-dire à repérer d'autres propriétés intuitives des contrefactuels. Une théorie correcte des contrefactuels devrait rendre compte de ces propriétés, en accordant un sens précis à ces intuitions floues, ou sinon les dissiper en les montrant illusoires.

Le premier point important a été constaté par Goodman : « la distinction entre les 'principes de connexion' [notion de conséquence] et les conditions pertinentes [prémisses] est imprécise et arbitraire ; les 'principes de connexion' [notion de conséquence] pourraient être ajoutés aux conditions pertinentes [prémisses], de sorte que la relation entre d'une part la conjonction de l'antécédent avec ces conditions et d'autre part le conséquent soit une question logique. »²⁹

C'est d'une telle évidence à ne pas appeler de commentaire supplémentaire. *Il n'y a pas de frontière stable entre les prémisses et la notion de conséquence qui contribuent à une analyse d'un contrefactuel.*

Il convient de remarquer le profit que l'on tire souvent du caractère imprécis et arbitraire de la frontière entre prémisses et la notion de conséquence : toutes les analyses de la littérature, à l'exception éventuelle de celle de Goodman qui n'est pas très claire sur ce point, imposent leur propre frontière *fixe* entre prémisses et notion de conséquence. Lewis, Stalnaker³⁰ et Mares & Fuhrmann posent des notions de conséquences faibles et logiques, pour permettre l'analyse d'autant des contrefactuels que possible avec une même relation de similitude. Du reste, Lowe prend le contre-pied : pour lui la notion de conséquence est très

²⁹ *Op cit.*, p 17.

³⁰ Stalnaker indique à certains endroits qu'il emploie sous le nom « monde possible » des alternatives disjointes et « présentes » dans un contexte donné (par exemple, 1984, p58, 1986, p118 *sq.*), ce qui impliquerait que la notion de conséquence correspond aux vérités communes à toutes ces alternatives, lesquelles peuvent changer entre contextes. Cependant, dans la mesure où Stalnaker ne thématise pas le rapport entre les différentes alternatives relatives aux contextes différents (et notamment les rapports entre les contextes où différents énoncés sont *en jeu*, voir ici §4.2.4), et *a fortiori* la question de la frontière imprécise et arbitraire, il n'a pas proprement pris en charge cet aspect de la question.

forte, de sorte qu'il n'y ait pas de prémisse. De même pour les probabilistes tel Adams et Edgington dans leur propre cadre : ils ne mettent en jeu que des probabilités conditionnées qui, en tant que structures de la fonction probabiliste plutôt que valeurs, sont de l'ordre de la conséquence. Dans la mesure où aucun de ces théoriciens ne s'efforce de comprendre ni de rendre compte du caractère imprécis de la frontière entre prémisses et la notion de conséquence, dans la mesure où ils rasent cette indétermination subtile pour les besoins de leurs propres théories, ils montrent un manque de sensibilité envers cette propriété des contrefactuels. Le caractère arbitraire, imprécis, voire même vague de la frontière entre prémisses et la notion de conséquence est un phénomène qui *doit être pris en compte* par une théorie des contrefactuels ; sur ce front, les théories existantes laissent à désirer.

4.2.3 Indépendance et pertinence

Où l'on met en évidence premièrement le fait que non seulement la prémisse d'un contrefactuel est indépendante de son antécédent, mais qu'elle l'est même en présence du conséquent, et deuxièmement le fait que l'antécédent d'un contrefactuel est normalement pertinent pour son conséquent. Il découle de ce deuxième fait une distinction entre les semifactuels non indépendants, dont l'antécédent est pertinent au conséquent, et les semifactuels indépendants, où il ne l'est pas. Cette distinction manque formellement aux théories proposées jusqu'à présent ; en outre, pour rendre compte des phénomènes qui seraient autrement expliqués à l'aide de cette distinction, les théoriciens ont dû recourir à la pragmatique.

Intuitivement, il y a trois relations entre les prémisses, l'antécédent et le conséquent d'un contrefactuel :

1. La valeur de vérité d'une prémisse n'est pas « affectée » par la valeur de vérité de l'antécédent du contrefactuel. On dira qu'il manifeste un certain degré « d'indépendance » par rapport à l'antécédent.
2. En outre, ce degré « d'indépendance » n'est pas mis en question par la mise en jeu du conséquent. On dira que la prémisse connaît un degré « d'indépendance » de l'antécédent « compte tenu » du conséquent.
3. Par ailleurs, les prémisses ne devraient pas permettre d'impliquer le conséquent indépendamment de l'antécédent. On dira que l'antécédent devrait être « pertinent » pour le conséquent.

Les deux premiers points seront mis en évidence et approfondis premièrement ; le dernier point sera considéré ensuite.

Indépendance et co-indépendance

Le genre de phénomène dont il sera question ici est bien illustré par un problème que se pose Goodman³¹. Il constate que, *du point de vue logique*, il est également valide d'affirmer « si cette allumette avait été grattée, alors elle s'enflammerait », en s'appuyant sur la prémisse « elle est sèche », que d'affirmer « si cette allumette avait été grattée, elle ne serait pas sèche », en s'appuyant sur la prémisse « elle ne s'enflamme pas ». Car, en écrivant A pour « cette allumette a été grattée », B pour « elle s'enflamme » et G pour « elle est sèche », le premier contrefactuel s'appuie sur la conséquence $A \& G \rightarrow B$, alors que le second s'appuie sur la conséquence $A \& \text{non } B \rightarrow \text{non } G$, et ces deux formules sont logiquement équivalentes. Face à cet apparent problème, on a tendance à affirmer que la différence tient au fait que G ne dépend pas de A ou de B et que, en revanche, c'est B qui dépend de A et de G . Il s'agit ici d'interroger l'intuition quotidienne à l'égard de cette « indépendance » et cette « dépendance ».

Tout d'abord, l'idée d'un rapport de l'ordre de « l'indépendance » des prémisses d'un contrefactuel à l'égard de son antécédent est assez naturelle et connaît d'ailleurs une longue histoire : la conclusion que Goodman lui-même tire de son exemple est que les prémisses d'un contrefactuel doivent être « co-tenables » avec l'antécédent, c'est-à-dire que « ce n'est pas le cas que S [l'ensemble de prémisses] ne serait pas vrai si A [l'antécédent] l'était. »³² L'idée générale (un peu plus audacieuse, mais également intuitive, que la formulation soigneuse goodmanienne) est apparemment que le changement (contrefactuel) de la valeur de vérité de A (l'antécédent) n'affecte pas la valeur de vérité de G (la prémisse)³³.

Cette relation de « ne pas être affecté » par la valeur de vérité d'un autre énoncé est apparemment complexe : en particulier, *le fait pour deux énoncés d'entrer dans une telle relation semble pouvoir être touché par le fait que d'autres énoncés sont en jeu*. On dirait bien par exemple que « l'allumette est grattée » et « l'allumette est sèche » sont des énoncés tels que la valeur de vérité de l'un n'a pas d'effet sur celle de l'autre, mais lorsque s'ajoute le fait qu'il pleut, cette

³¹Goodman (1954, p14).

³²Goodman, *op cit.*, p15. Dans le même passage Goodman affirme, relatif à l'exemple présenté, que « le problème résulte de la présence dans S [l'ensemble des prémisses] d'un énoncé vrai qui, quoique compatible avec A [l'antécédent] ne serait pas vrai si A l'était. »

³³Sinon, crie-t-elle l'intuition quotidienne, comment pourrait-on compter sur C comme prémisse ?

situation est bouleversée, dans la mesure où le fait de gratter l'allumette (et par conséquent de la sortir de la boîte d'allumettes) nuit à sa sécheresse (en raison de la pluie). Il en découle les deux conclusions de ce paragraphe à l'égard des relations entre les prémisses, l'antécédent, et le conséquent d'un contrefactuel.

La première a trait au genre « d'indépendance » intuitivement en jeu lorsqu'il s'agit de plus de deux énoncés : il y a une différence entre les cas où les énoncés sont « indépendants » deux à deux – pour chaque paire d'énoncés, la valeur de vérité de l'un n'est pas affectée par celle de l'autre – et les cas où, de surcroît, *chacun des énoncés est indépendant de tous les autres énoncés pris ensemble* – la valeur de vérité de *chaque énoncé* n'est pas affectée par les valeurs de vérité de *tous les autres énoncés*. L'exemple de la pluie et de l'allumette illustre cette subtilité. En écrivant D pour « il pleut », A et D sont bien « indépendants » l'un relativement à l'autre (le fait de gratter l'allumette n'affecte pas le fait qu'il pleuve, ni l'inverse), G et D sont « indépendants » l'un relativement à l'autre (la sécheresse de l'allumette, qui est dans sa boîte, n'est pas affectée par la pluie, et inversement) et *idem* pour A et G ; or, comme on l'a déjà remarqué, les valeurs de *deux* d'entre eux (D et A) pourraient affecter celle du troisième (G). Chacune des paires semblent ne pas affecter l'un l'autre prise « isolément », pour ainsi dire, mais parmi les trois énoncés, il y a d'éventuels effets des valeurs de vérités *des uns* par rapport *aux autres*. On appellera *ensemble co-indépendant* un ensemble d'énoncés dont, pour *chaque* membre, il n'y a aucun effet des valeurs des vérités de *tous* les autres sur sa valeur de vérité. En pratique, l'ensemble constitué par les prémisses envisagées pour un contrefactuel et son antécédent est le plus souvent un *ensemble co-indépendant* : pour le contrefactuel « si l'allumette avait été grattée, alors elle s'enflammerait », on prend coutumièrement des prémisses telles « il y a assez d'oxygène », « elle est bien faite », « elle est sèche », lesquelles constituent avec l'antécédent un ensemble co-indépendant, plutôt que des ensembles de prémisses qui comportent des dépendances internes (sans ou avec la mise en jeu de l'antécédent), telle la paire « elle est sèche » et « il pleut », ou le triplet « elle est sèche », « elle est bien faite », « elle est sèche et bien faite ».

Le deuxième point concerne le rapport entre d'une part l'ensemble contenant les prémisses et l'antécédent, et d'autre part le conséquent du contrefactuel : si la mise en jeu d'un autre énoncé peut briser « l'indépendance » et *a fortiori* la « co-indépendance » d'un ensemble d'énoncés, alors est-il légitime d'employer un ensemble de prémisses tel que l'ensemble constitué de ces prémisses et l'antécédent *n'est plus* « co-indépendant » une fois mis en jeu le conséquent du contrefactuel ?

Il s'avère que la réponse est négative ; un exemple permettra de mieux apercevoir pourquoi³⁴.

(E1) « Si la ville de New York était dans l'état de Georgie, alors la ville de New York serait dans le Sud »

(E2) « Si la Georgie incluait la ville de New York, alors Georgie ne serait pas complètement dans le Sud »

La *structure* de cet exemple ressemble à celle de l'exemple précédent : en écrivant A pour « la ville de New York était dans l'état de Georgie », B pour « la ville de New York est dans le Sud », et G pour « la Georgie est complètement dans le Sud », les deux contrefactuels mettent en jeu les formules $A \& G \rightarrow B$ et $A \& \text{non } B \rightarrow \text{non } G$ respectivement, qui sont logiquement équivalentes. En outre, les trois énoncés sont deux à deux « indépendants », mais, en raison de la formule qui les relie, ils ne constituent pas un ensemble co-indépendant. Cet exemple admet cependant plusieurs « lectures » (selon les contextes différents) qui accordent des valeurs de vérités différentes aux contrefactuels. Une lecture accorde la vérité à un des contrefactuels (disons (E1)), et la fausseté à l'autre ; dans ce cas, on affirme que A et G sont bien « indépendants » et que B ne pourrait pas toucher à cette indépendance dans la mesure où sa valeur de vérité « dépend » de celles de A et G . Le fait pour New York d'être dans la Georgie n'a pas « d'effet » sur le fait que Georgie est dans le Sud, *même si* l'on met en jeu le fait que New York n'est pas dans le Sud, car ce dernier fait dépend, selon cette lecture, de la situation de Georgie et le fait ou non pour New York d'être dedans. D'un autre côté, il y a une lecture où l'on ne voit pas comment trancher entre les contrefactuels, en décidant que les deux sont indéterminés ou tout simplement faux³⁵ ; dans ce cas, on ne nie pas que les deux énoncés A et G soient « indépendants » *pris isolément*, ni d'ailleurs que ce soit le cas pour A et B , mais on est sensible au fait que la mise en jeu du troisième énoncé *compromet* cette « indépendance ». Le fait pour New York d'être dans Georgie n'a pas « d'effet » sur le fait que Georgie soit dans le Sud si l'on considère ces deux énoncés seulement, mais dès que le fait pour New York d'être ou non dans le Sud entre en question, il y a des « effets » possibles (puisque New York n'est pas dans le Sud, s'il était dans Georgie, alors Georgie ne serait pas complètement dans le Sud). Pour valider le contrefactuel, il faut non seulement

³⁴Il s'agit encore d'un exemple de Goodman, *op cit.*, p15, qu'il emploie de manière différente.

³⁵Il y avait un débat autour de contrefactuels de ce genre entre Lewis (1973b, §3.4) et Stalnaker (1980), dans lequel Lewis soutenait que les deux contrefactuels sont faux, alors que Stalnaker soutenait qu'ils sont les deux indéterminés. Ce débat est tangentiel à la discussion menée ici.

que les prémisses et l'antécédent constituent un ensemble co-indépendant, mais que *cette co-indépendance ne soit pas touchée par la mise en jeu du conséquent*. C'est ce qu'affirme la première lecture des contrefactuels – puisque le conséquent « dépend » de l'antécédent et des prémisses, il ne peut pas « affecter » leur (co-)indépendance; c'est ce que nie la dernière lecture – le conséquent peut « affecter » la (co-)indépendance des autres énoncés dans la mesure où, alors qu'il entre dans une relation non triviale avec eux (en l'occurrence, $A \& G \rightarrow B$), sa valeur de vérité n'est pas « assujettie » à les leurs. On dira que *les prémisses et l'antécédent forment un ensemble qui est co-indépendant compte tenu du conséquent*.

On est actuellement, rappelle-t-on, dans une étape préliminaire du travail où il s'agit de récolter des phénomènes relatifs aux contrefactuels, aussi vagues ou éphémères fussent-ils; dans le chapitre suivant on développera une analyse qui rend compte de ces aspects et qui donne notamment un sens clair et précis à ces notions floues « d'indépendance » ou de « dépendance ». Incidemment, la notion d'indépendance, ou des notions apparentées, n'est pas inconnue dans la littérature sur les contrefactuels. Non seulement Goodman a un concept semblable (la « co-tenabilité »), mais l'analyse de Lewis (et de Stalnaker) comporte des aspects semblables. Une prémisses P pour un contrefactuel $A \Rightarrow B$ évalué à un monde i est l'ensemble de mondes qui sont plus proches ou également proches de i que le(s) monde(s) le(s) plus proche(s) de i où A est vrai (§4.2.1). Or, puisque P est vrai dans le(s) monde(s) le(s) plus proche(s) où A est vrai, le contrefactuel $A \Rightarrow P$ est vrai³⁶. Ce contrefactuel affirme que si A était le cas, P serait (toujours) le cas, ce qui est une façon primitive et faible de dire que la valeur de vérité de A n'affecte pas celle de P . Une prémisses d'un contrefactuel dans l'analyse lewisienne n'est donc pas « affectée » par l'antécédent; elle lui est « indépendant ».

Pertinence de l'antécédent

Caractéristique intuitivement perceptible mais difficile à préciser, il semble que, pour des contrefactuels « typiques » ou « propres », l'antécédent ait un « effet » ou soit « pertinent » relatif au conséquent. Lorsque cette « pertinence » manque, l'intuition à l'égard du contrefactuel est trébuchante et complexe; pour dire le moins, le contrefactuel est problématique. En approfondissant ce genre de « pertinence » sur plusieurs exemples, on trouvera une distinction forte entre les contrefactuels où elle subsiste et ceux où elle est absente. Ce genre de « per-

³⁶Voir Lewis (1973b, §2.6) ou Lewis (1973c, §2.2) pour les détails.

tinence » et cette distinction sont donc des phénomènes dont une théorie des contrefactuels devrait rendre compte. Or il s'avérera, dans un deuxième temps, que les approches de cette question dans la littérature, si elles ne méprisent pas le phénomène, en rendent compte seulement grâce à un fort appui sur la pragmatique.

Voici un problème pour une analyse naïve des contrefactuels. Pour n'importe quel P et A , on a $A \& P \rightarrow P$ par la logique. Or, par l'application naïve de la forme d'analyse (**AGC**), il en découle que, pour n'importe quel énoncé P qui est « indépendant » de A (dans un sens approprié), $A \Rightarrow P$ si et seulement si P est vrai³⁷. Or, cette conclusion entre en tension avec l'intuition quotidienne à l'égard des contrefactuels, quelle que soit la valeur de vérité de A . Grâce au rapport entre contrefactuels et fiction (§4.1.2), on pourrait d'employer les vérités fictionnelles à titre d'exemple. Pour donner un exemple où A est faux, prendre pour A , « les *Aventures de Sherlock Holmes* est cru être vrai » et pour P , « Jacques Chirac est président de la France en 2004 » : par une analyse naïve $A \Rightarrow P$, mais dans les *Aventures de Sherlock Holmes*, Jacques Chirac est-il le président de France en 2004 ? Ce n'est cependant pas seulement l'intuition relative à la fiction qui est froissée : en changeant le A , on obtient le contrefactuel « si l'allumette était grattée, alors Jacques Chirac serait président de la France en 2004 ». De l'autre côté, pour illustrer les cas où A est vrai et le contrefactuel est contraire à l'intuition, on pourrait prendre les cas de l'application de l'analyse de la vérité fictionnelle à des textes non fictionnels³⁸ : telle biographie de Napoléon est (crue) vraie (A), et l'hiver de 2004 est particulièrement froid (P), mais on n'est pas enclin à admettre que « dans telle biographie de Napoléon, l'hiver de 2004 est particulièrement froid ». De manière semblable, le contrefactuel « si Napoléon avait gagné à Austerlitz, alors l'hiver de 2004 serait particulièrement froid » fait trébucher l'intuition.

Deux intuitions relatives à ces exemples se confondent. D'un côté, s'il y a une intuition qui pousse à évaluer ces contrefactuels comme vrais, elle semble tenir à la vérité du conséquent et à elle seule, indépendamment de l'antécédent.

³⁷Ce constat, et les exemples suivants qui s'y reposent, ne s'applique qu'aux analyses des contrefactuels qui emploient une notion de conséquence qui valide $A \& P \rightarrow P$, dont l'implication classique, mais non, par exemple, l'implication pertinente. Voir la note 39.

³⁸On pourrait douter de l'application de cette analyse aux textes non fictionnels. On poserait ainsi à côté de l'opérateur « dans le texte de fiction f , ... » un opérateur « dans le texte de non fiction n , ... », de sorte que les deux s'opposent à l'opérateur « canonique » « dans le monde hors texte, ... » (qui est équivalent à l'absence d'opérateur). Non seulement cette distinction semble-t-il peu justifiée, mais de plus, une analyse de l'opérateur « dans un texte de non fiction n , ... » devra surmonter la même difficulté de l'excès de « vérités selon le texte » que l'analyse de l'opérateur fictionnel, c'est-à-dire le même problème de *verum ad quodlibet*.

En cela, la situation rappelle le vieux problème de *verum sequitur ad quodlibet* touchant à l'implication matérielle et au fait que ce genre d'implication est vraie si le conséquent est vrai. De l'autre côté, il y a sans doute une intuition inverse qui pousse à rejeter ce genre de contrefactuel. L'exemple de la fiction est particulièrement clair là-dessus : on voudrait réserver la notion de « vérité fictionnelle », ou « vérité selon un texte », pour les vérités *autorisées* d'une manière ou d'une autre par le texte, plutôt que de permettre de surcroît tous les énoncés vrais du monde réel mais qui ne sont pas du tout touchés, de quelle manière que ce soit, par les énoncés du texte. De manière générale, on veut une *pertinence* de l'antécédent par rapport au conséquent.

Il y a donc un certain nombre de semifactuels – contrefactuels avec conséquent vrai – où il n'y a apparemment pas de pertinence de l'antécédent par rapport au conséquent, et où l'intuition trébuche. Or, il n'est pas vrai que tout semifactuel soit de ce type ; il n'est pas vrai que tout semifactuel connaisse ce phénomène de *verum ad quodlibet* (si l'on veut l'affirmer comme vrai, c'est entièrement en raison de la vérité du conséquent). Considérons par exemple le semifactuel « (même) si l'allumette avait été grattée, elle ne s'enflammerait (toujours) pas ». Si ce semifactuel est vrai, on ne penchera pas à dire que c'est en raison d'une indépendance entre l'antécédent et le conséquent : on ne dira certainement pas qu'elle ne s'enflamme pas parce que la gratter ne fait aucune différence à ce qu'elle s'enflamme ou non. Donc il n'est pas question de *verum ad quodlibet* ici. Au contraire, on le traitera de même façon que des contrefactuels ordinaires : on cherchera une raison pour la vérité, laquelle jouera un rôle analogue à celui joué par les prémisses dans l'analyse ordinaire des contrefactuels. Par exemple, on s'attendrait à une affirmation telle « si l'allumette avait été grattée, elle ne s'enflammerait pas, car il y avait trop de vent ». La vérité du semifactuel s'explique ainsi : « l'allumette a été grattée & il y a trop de vent \rightarrow elle ne s'enflamme pas ».

Par conséquent, une distinction s'impose entre deux sortes de semifactuels : les semifactuels *non indépendants* où le conséquent n'est pas indépendant de l'antécédent, et les semifactuels *indépendants*, où il l'est. Un exemple du premier type est « si l'allumette avait été grattée, elle ne s'enflammerait pas (car il y avait trop de vent) » ; un exemple du dernier type est « si l'allumette était grattée, Jacques Chirac serait le président de la France (car les deux choses n'ont rien à voir) ». Le phénomène de *verum ad quodlibet* ne se produit que dans les cas de semifactuels indépendants.

Cette distinction renvoie à un phénomène concret : la différence entre les semifactuels où l'intuition hésite sur leur valeur de vérité en raison d'un manque

de pertinence de l'antécédent pour le conséquent (les cas de *verum ad quodlibet*) et les semifactuels où il n'y a pas de telle hésitation. Ce phénomène, cette différence même, compte comme un défi pour une éventuelle théorie des contre-factuels. Que l'on souligne que la question *n'est pas* celle de la *vérité* de ces semifactuels : il ne s'agit pas de la valeur de vérité on devrait ou voudrait accorder aux semifactuels indépendants. Il est plutôt question de leur *identification* : il faut qu'une théorie puisse *reconnaître un semifactuel comme indépendant et non indépendant* (et éventuellement, laisser au théoricien la possibilité de décider de la vérité des semifactuels indépendants à son gré). Or, presque toute analyse proposée jusqu'à présent manque la distinction entre semifactuel indépendant et semifactuel non indépendant, et donc n'arrive pas à rendre compte de cette différence³⁹. Il n'apparaît explicitement ni chez Goodman, ni chez Lewis ou Stalnaker, ni chez Adams ou Edgington.

En ce qui concerne l'analyse en termes d'une relation de similitude entre mondes possibles, on connaît bien que, dans le cas où l'antécédent est vrai, l'analyse réduit le contrefactuel à l'implication matérielle⁴⁰. Donc, dans l'exemple « si Napoléon avait gagné à Austerlitz, alors l'hiver de 2004 serait particulièrement froid », puisque le monde le plus proche du nôtre où Napoléon gagna est bien le nôtre, et puisque l'hiver de 2004 était froid dans notre monde, ce contrefactuel est évalué comme carrément vrai. De même dans les cas d'antécédent faux. Reprenant l'exemple « si l'allumette était grattée, alors Jacques Chirac serait

³⁹ L'importance de la *pertinence* de l'antécédent d'une implication pour son conséquent que est en jeu ici connaît une longue histoire en logique. Au temps moderne, cette tradition prend la forme de la dite « logique de la pertinence », introduite dans les travaux fondateurs de Ackermann, Anderson et Belnap (voir Anderson and Belnap (1975) pour des articles et d'autres références), et proposant une « implication pertinente » qui relie un antécédent seulement avec les conséquents pour lesquels il est pertinent. Les artisans de la logique de la pertinence n'ont pas tardé à constater que la question de la pertinence se pose également pour les contrefactuels (Dunn, 1986), et à proposer des analyses des contrefactuels qui ressemblent à l'analyse classique en termes de mondes possibles à ceci près que l'implication matérielle est remplacée par l'implication pertinente (Mares and Fuhrmann (1995), Mares (2004) et §4.2.1 ci-dessus). Ces analyses n'évaluent pas les semifactuels indépendants comme vrais ; pourtant, elles les évaluent comme simplement faux. Pour cette raison, il n'est pas clair qu'elles permettent de *reconnaître* la différence entre les semifactuels indépendants et les semifactuels non indépendants, comme l'est exigée ici ; en tout cas, si elles peuvent, les théoriciens de la pertinence ne thématisent pas cette possibilité (au mieux, Mares and Fuhrmann (1995, pp 646-7) suggèrent une lecture différente des semifactuels indépendants, ce qui suppose la différence plutôt que de la repérer). En revanche, un exemple plus notable est Kvart (1986, par exemple Ch. 2), dont le cadre et le questionnement diffèrent de ceux en vigueur ici, mais dont la notion de « semifactuel non pertinent » ressemble à celle de semifactuel indépendant.

⁴⁰Le monde plus proche du monde actuel est le monde actuel (seulement) ; Lewis (1973b, §1.7).

président de la France en 2004 », il semble y avoir des mondes où l'allumette est grattée et Jacques Chirac est président qui sont beaucoup plus proches du nôtre que tout monde où elle est grattée mais il n'est pas président. De telle sorte que le contrefactuel est évalué comme carrément vrai. On ne voit aucune différence avec l'évaluation des contrefactuels « typiques » et *a fortiori* avec des semifactuels non indépendants, qui se réfèrent également à la vérité du conséquent dans les mondes les plus proches.

Dans le cas de l'analyse probabiliste, les choses sont aussi claires. La théorie de la probabilité a sa propre notion d'indépendance, qui conçoit l'indépendance de P à l'égard de A comme l'égalité des probabilités $p(P/A)$ et $p(P)$. Pour chacun des contrefactuels donnés comme exemple ci-dessus, sa probabilité est égale à la probabilité de son conséquent ; et puisque le conséquent a une grande probabilité (il est « affirmable » comme vrai), le contrefactuel l'a aussi (il est « affirmable » comme vrai). Encore une fois, l'hésitation de l'intuition qui caractérise ces cas et les distingue des semifactuels non indépendants n'est pas manifeste dans la perspective probabiliste.

Pour autant que les théoriciens des contrefactuels se soient attardés sur la question des semifactuels indépendants et sur le phénomène de *verum ad quodlibet* qui y apparaît, ils ont adopté envers eux trois genres d'approche. La première consiste à les dévaloriser, au motif du caractère inouï de ces constructions contrefactuelles. C'est la position de Lewis par rapport aux cas où l'antécédent et le conséquent sont vrais : « l'étrangeté n'est pas la fausseté »⁴¹. Certes, mais cela n'implique pas non plus que l'étrangeté soit la vérité, comme l'affirme son analyse.

Les deux autres approches ont en commun un appui important sur des aspects *pragmatiques*, quoique des aspects pragmatiques apparemment différents. L'une d'entre elles repose sur les manipulations de la part non formelle nécessaire à toute analyse des contrefactuels (la notion de conséquence et les prémisses, la relation de similitude, la fonction probabiliste), qui a été reconnue dans §4.2.1 comme un facteur ressortissant à la pragmatique. On appellera cet aspect de la pragmatique la *pragmatique engagée*, parce qu'elle « engage » directement dans la théorie. Par contraste, l'autre approche met en œuvre d'autres aspects relevant de la pragmatique, telle le vague ou l'indétermination des termes ou des concepts, qui « planent au-dessus » de la logique des contrefactuels (règles logiques, système de mondes possibles) sans engager avec elle : la détermination de ces aspects, alors qu'elle est importante pour la question de *l'assertabilité* des contrefactuels, ne touche aucunement à leur *vérité*. On appellera ces aspects de

⁴¹Lewis (1973b, p28).

la pragmatique, qui se tiennent *entre* les faits quotidiens du langage (les « affirmations ») et un noyau logique ou sémantique (les « vérités »), les aspects de la *couche pragmatique*.

L'approche qui a recours à la pragmatique engagée soutient, face à l'exemple « si l'allumette était grattée, alors Jacques Chirac serait président de la France en 2004 », que la relation de similitude en œuvre dans le contexte actuel est indifférente à l'identité du président de la France⁴². Pour chaque monde où l'allumette est grattée et le président est Jacques Chirac, il y en a un au moins également proche du nôtre où elle est grattée et il ne l'est pas, de telle sorte que ni le contrefactuel ($A \Rightarrow B$) ni son « contraire » ($A \Rightarrow \neg B$) soient vrais. La même astuce pourrait être employée dans le cas où l'antécédent est vrai (« si Napoléon avait gagné à Austerlitz, alors l'hiver de 2004 serait particulièrement froid »), en permettant qu'il y ait des mondes inactuels qui sont aussi proches du monde actuel que lui-même⁴³. Par exemple, il y a des mondes aussi proches du monde actuel que lui-même où l'hiver de 2004 n'est pas froid.⁴⁴

Par contraste, l'autre approche de la question de *verum ad quodlibet* et des semifactuels indépendants met en œuvre la *couche pragmatique*. Stalnaker soutient que l'aspect de la pragmatique engagée – la relation de similitude par exemple – est fourni *de façon indéterminée* : la pragmatique ne délimite qu'un ensemble de relations de similitudes, et le contrefactuel, pour être vrai, doit être évalué comme vrai selon chacune⁴⁵. Dans les cas où l'antécédent est faux, tel « si l'allumette était grattée, alors Jacques Chirac serait président de la France en 2004 », la valeur de vérité du contrefactuel est indéterminée en raison de l'indétermination de la relation de similitude : on ne sait pas si Jacques Chirac est président dans le(s) monde(s) le(s) plus proche où l'allumette est grattée ou non. En revanche, cette approche est moins satisfaisante dans les cas où l'antécédent est vrai : quitte à renoncer à la supposition que tout monde est

⁴²Lewis suggère ce genre de réponse dans 1978a, pp465-7.

⁴³Lewis considère cette option de plus près dans 1973b, §1.7.

⁴⁴Par parenthèse, cette manière d'aborder les exemples de *verum ad quodlibet* ne marche pas. Car, si, comme le recommande Lewis (1973b, §1.5.), on interprète une situation où P est vrai dans certains des mondes les plus proches où A est vrai et faux dans d'autres comme un contrefactuel « pourrait » – « si A était vrai, alors P pourrait être vrai » –, on échange les contrefactuels ordinaires contre des contrefactuels « pourrait ». On n'a plus « si Napoléon avait gagné à Austerlitz, alors l'hiver de 2004 serait particulièrement froid », mais plutôt « si Napoléon avait gagné à Austerlitz, alors l'hiver de 2004 pourrait être particulièrement froid » ; on n'a plus « si l'allumette était grattée, alors Jacques Chirac serait président de la France en 2004 » mais plutôt « si l'allumette était grattée, alors Jacques Chirac pourrait être président de la France en 2004 ». Ce qui ne soulage pas les inquiétudes. Néanmoins, ce qui importe pour la discussion présente n'est pas l'efficacité de ces approches, mais leur appui sur la pragmatique.

⁴⁵Stalnaker (1980, p89 *sq.*).

le monde le plus proche de lui-même, on ne sait pas dire que le monde pertinent pour l'évaluation du contrefactuel est indéterminé, car c'est bien le monde actuel.

Dans tous les cas, c'est *en s'appuyant sur des aspects pragmatiques*, tantôt sous forme de relations de similitude habilement sélectionnées, tantôt sous forme d'heureuses indéterminations dans ces relations, *que les théories espèrent rendre compte de l'invraisemblance a priori de leurs analyses*. Dans la mesure où, comme on verra prochainement, l'aspect pragmatique n'est pas compris, la pertinence de l'antécédent au conséquent, et la complexité des cas où elle est absente, *n'est pas vraiment comprise non plus*.

4.2.4 Le *hors sujet*

Où la notion de hors sujet d'un énoncé par rapport à un contexte est introduite. Le hors sujet n'est pas la même chose que l'indépendance par rapport au contrefactuel en question ; en fait, il est plutôt relié à l'idée de l'absence par rapport au contexte.

Si, à la fin d'un discours où l'on soutient qu'une allumette est bien faite, on affirme le contrefactuel « si l'allumette avait été grattée, alors elle s'enflammerait », l'énoncé « l'allumette est bien faite » figure comme prémisses dans le contexte de l'affirmation de ce contrefactuel. On a déjà constaté que, en tant que prémisses, cet énoncé est indépendant de l'antécédent du contrefactuel. Or, comme déjà remarqué, l'énoncé « Jacques Chirac est président de la France » est lui aussi indépendant de l'antécédent. Néanmoins, on incline à accorder un statut différent à « l'allumette est bien faite » qu'à « Jacques Chirac est président de la France ». Le premier prend place dans la discussion, le dernier n'a rien à y faire. Le premier est *présent* dans le contexte, le dernier y est dans un certain sens *absent*. Le premier est *en jeu*, le dernier est *hors sujet*.

Cette notion de *hors sujet*, reconnue d'ailleurs d'importante dans la considération de la fiction (§§1.4.1, 2.2, 2.5.3, 3.1.1), sera également d'une importance capitale pour les contrefactuels. Alors qu'elle est difficile et éphémère, pour la raison simple que l'acte même par lequel on exhibe quelque chose de *hors sujet* lui ôte son caractère de *hors sujet*, on peut en constater quelques propriétés, dont deux sont particulièrement importantes relatives à la question des contrefactuels. La première est que le *hors sujet* est *distinct de l'indépendance relative au contrefactuel*, c'est-à-dire, relative à son antécédent et à son conséquent. La deuxième – qui rejoint les analyses de la première partie et inspirera l'analyse des contrefactuels qui sera développée dans la suite – met en évidence un rapprochement entre le *hors sujet* et l'absence par rapport à l'algèbre, dans la conception

algébrique de contrefactuels (§4.2.1).

En ce qui concerne le rapport entre l'indépendance et le *hors sujet*, le fait que, dans l'exemple donné ci-dessus, on peut priver la présidence de Jacques Chirac de son statut de *hors sujet*, en le mentionnant, sans toucher à son indépendance relative au contrefactuel (c'est-à-dire, à son antécédent et son conséquent) montre que l'indépendance d'un énoncé n'implique pas qu'il soit *hors sujet*. A l'inverse, le *hors sujet* n'implique pas l'indépendance. Pour revenir encore à l'exemple précédent, des lois complexes de la mécanique quantique sont *hors sujet*, alors qu'elles ne sont pas nécessairement indépendantes relativement au contrefactuel « si l'allumette avait été grattée, alors elle s'enflammerait ». Car il semble possible *en droit*, quoique un peu excessif, de préciser les lois de la physique impliquées dans le fait de prendre feu suite à l'action de la gratter au niveau de détail impliqué dans la théorie de la mécanique quantique, de telle sorte que la loi de la mécanique quantique figure comme prémisse dans l'analyse du contrefactuel, et donc, pour dire le moins, comme non indépendant relativement au conséquent. Pour *hors sujet* qu'elle soit, la loi de la mécanique quantique n'est pas indépendant du contrefactuel, dans la mesure où, si elle était *en scène*, elle ne lui serait pas (nécessairement) indépendant. Par conséquent, le *hors sujet* n'implique pas l'indépendance à l'égard du contrefactuel.

On arrive à mieux comprendre ce point en recourant à la conception algébrique de la notion de conséquence et des prémisses. Toute analyse du contrefactuel implique quelque chose de l'ordre d'une régularité physique, reliant l'action de gratter l'allumette à son inflammation. Le cas le plus simple est celui où cette régularité est employée comme conséquence – c'est-à-dire comme aspect de la *structure* de l'algèbre de Boole (§4.2.1) – qui relie « elle est grattée » et « elle s'enflamme ». Alors qu'il y a tout un faisceau de relations et lois physiques, de complexités, de réserves et de détails qui soutiennent une telle régularité en général, *tout ce qui est présent dans la structure de l'algèbre est ce rapport entre ces deux énoncés*. Les lois physiques les plus détaillées y sont absentes. En outre, les énoncés relatifs à la présidence de la France sont également absents de cette algèbre. Dans la mesure où seulement ce qui est présent dans cette algèbre contribue à l'évaluation du contrefactuel, on saurait rapprocher l'*absence* de ces aspects de leur caractère de *hors sujet*.⁴⁶

⁴⁶La différence entre l'exemple de la loi physique et celui de la présidence de Jacques Chirac tient à la « raison » de cette absence : dans le cas des lois physiques, ce qui est *hors sujet* est « derrière » la structure, alors que dans le cas de Jacques Chirac, ce qui est *hors sujet* est « à côté » de la structure. Cette différence rappelle celle entre la profondeur et la largeur discutée dans §2.3.1. Il n'y a pas de place ici pour s'attarder sur cette distinction et ses subtilités techniques (voir cependant la note 45 de §2.3.1).

Le *hors sujet* a donc trait, comme dans les considérations de la première partie (§§2.2, 2.5.3, 3.1.1), à l'*absence*, en l'occurrence l'absence de l'algèbre pertinente pour l'évaluation du contrefactuel. Il est également un phénomène dont une théorie de contrefactuels devrait rendre compte. Pourtant, la notion de *hors sujet* échappe aux analyses courantes, comme l'est témoigné par le fait qu'elles sont toutes globales, dans le sens définis dans §4.2.1. Par exemple, pour n'importe quel énoncé relatif aux lois de la physique, ou à la présidence de la France, tout contrefactuel figurant cet énoncé est évalué dans tous les contextes, même ceux où il est apparemment *hors sujet*.

4.3 Troisième tour : la question « difficile » des contrefactuels

4.3.1 La question des contrefactuels : sémantique et pragmatique

Où est mise en question la distinction répandue entre la partie sémantique et la partie pragmatique des contrefactuels. La distinction est effectivement douteuse, dans la mesure où la frontière exacte entre ces deux parties est difficile à tracer. Plus grave, elle encourage des pratiques méthodologiques fâcheuses. À savoir, d'un côté, un abus de la pragmatique, consistant en l'habitude de référer les phénomènes non expliqués par les théories sémantiques à la pragmatique, et de l'autre côté, un dédain de la pragmatique, comme l'est témoigné par la tendance à se concentrer sur la partie sémantique au détriment de la partie pragmatique. Cette dernière conséquence est d'autant plus regrettable que la vraie question des contrefactuels – celle de leur rapport à la réalité – appartient à la pragmatique. Une théorie qui prend proprement en charge la pragmatique ne ressemblerait que peu aux théories proposées jusqu'ici : notamment, elle ne s'occuperait pas de la question des « règles logiques » pour les contrefactuels, et elle ne serait pas nécessairement globale, au sens de §4.2.1.

« Je me préoccupe principalement de ce qui a été appelé le *problème logique des conditionnels*, problème qui a été souvent ignoré ou écarté par des spécialistes des conditionnels et des contrefactuels. »⁴⁷ Avec ces mots Stalnaker ouvre son fameux article sur les contrefactuels. Ce problème, précise-t-il

⁴⁷Stalnaker (1968, p165).

par la suite, s'oppose au *problème pragmatique des contrefactuels*. Plus : une résolution du problème logique, en fournissant une théorie sémantique, permettra la distinction entre les deux problématiques. « La théorie sémantique que je soutiens aidera à éclairer le [problème pragmatique] en traçant la frontière entre les composants sémantique et pragmatique du concept. »⁴⁸

Et pourtant, il y a de bonnes raisons pour se méfier de cette distinction tranchante entre sémantique (ou « logique ») et pragmatique. Non seulement elle est effectivement difficile à tirer. Non seulement elle conduit à des pratiques douteuses du point de vue méthodologique, où l'on reporte à la pragmatique toutes les questions que la théorie dont on est partisan ne permet pas d'aborder. Mais, de plus grave, elle a pour conséquence un dédain de la pragmatique, lequel est d'autant plus regrettable que le problème le plus difficile, le plus intéressant et le plus important relatif aux contrefactuels appartient à ce qui est délaissé sous nom de la pragmatique. Il s'agit dans cette section d'ébranler cette distinction, en soutenant d'une part qu'elle est strictement intenable et d'autre part qu'elle est méthodologiquement fâcheuse, pour ensuite soutenir l'importance de la pragmatique et quelques conséquences de sa prise en charge.

Instabilité de la distinction entre sémantique et pragmatique

Tout d'abord, la manière de diviser les propriétés des contrefactuels, et par conséquent les problèmes qu'ils posent, en une partie sémantique ou logique et une partie pragmatique ne va absolument pas de soi. Le bref parcours de la littérature sur les contrefactuels effectué dans §4.2.1 suffit à mettre en évidence la diversité de théories, dont chacune trace *sa propre limite* entre le composant sémantique et le composant pragmatique ; dont chacune cerne, pour ainsi dire, *son propre* « *noyau sémantique* » dont elle entend rendre compte. Même parmi les théories qui s'expriment en termes de similitudes entre mondes, il y en a une diversité, préconisant des propriétés *différentes* des relations de similitude et donc, enfin, des lignes de partage *différentes* entre la partie sémantique et la partie pragmatique de la question⁴⁹.

En outre, il y a très peu de propriétés logiques ou sémantiques indubitables des contrefactuels qui permettraient de préférer l'un des prétendus noyaux sémantiques sur les autres. Pire : même les « règles logiques conventionnelles », lesquelles sont admises par la plupart de théoriciens comme un aspect important de la partie sémantique ou logique des contrefactuels, ne sont pas hors de doute⁵⁰. Je sors dans la rue ; je n'aime pas me mouiller. Il semble donc que

⁴⁸*Op cit.*, p166.

⁴⁹Voir Lewis (1981) pour un compte rendu de certains d'entre eux.

⁵⁰Une liste de base de telles règles pour la logique non-monotone se trouve dans l'Annexe

les contrefactuels « s'il pleuvait, j'apporterais une parapluie » et « s'il pleuvait et j'apportais une parapluie, je ne me mouillerais pas » sont admissibles ou vrais ; par contraste le contrefactuel « s'il pleuvait, je ne me mouillerais pas » est beaucoup moins évident. Pourtant, selon une forme d'argumentation validée par presque toute théorie des contrefactuels – appelée *Transitivité Cumulative* ou simplement *Coupure*⁵¹ – le dernier contrefactuel découle des deux premiers. Ou encore, « s'il pleuvait, je me mouillerais » et « s'il pleuvait, j'apporterais une parapluie » sont apparemment des contrefactuels vrais ou admissibles, alors que le contrefactuel « s'il pleuvait et j'apportais un parapluie, je me mouillerais » ne semble pas s'ensuivre. Cependant, cette forme d'argumentation, appelée *Monotonie Prudente*⁵², est elle aussi validée par presque toutes les théories des contrefactuels. Sans des formes d'argumentation ou des règles d'inférence qui sont validés universellement et indubitablement par les contrefactuels, on ne saurait pas proposer un noyau sémantique qui est la visée d'une théorie sémantique de contrefactuels, ni préférer un candidat de noyau sémantique sur un autre.⁵³

En somme, la pluralité de « noyaux sémantiques » préconisés, alliée à la difficulté de trouver des propriétés sémantiques indubitables permettant de préférer l'un sur l'autre, pourrait ébranler la thèse stalnakerienne d'un problème sémantique, consistant à rendre compte *du* (seul) noyau sémantique.

Conséquences méthodologiques fâcheuses

La difficulté de fixer une fois pour toute des propriétés d'un noyau sémantique n'est pas sans rapport au fait que les « composants » sémantiques et pragmatiques des contrefactuels sont tellement enchevêtrés qu'il est impossible de les aborder, et surtout de les « tester », séparément. Comme le souligne Lewis, « c'est la combinaison [de la théorie sémantique et la contribution pragmatique] qui pourrait être testée contre notre connaissance de contrefactuels, non pas l'analyse par elle-même. »⁵⁴ Puisqu'il n'y a aucune manière d'« accéder » à la partie sémantique en isolation, différentes prétendues propriétés sémantiques des contrefactuels pourraient être défendues en référant leurs contre-exemples respectifs à la pragmatique. Cette possibilité est révélateur des conséquences

A.1.

⁵¹Le système de Lewis (et *a fortiori* celle de Stalnaker) valide cette forme d'argumentation (Lewis, 1973b, p35). Pour plus de discussion de cette règle, voir Kraus et al. (1990, § 3.1) ou Makinson (2005, § 1.2).

⁵²Kraus et al. (1990, § 3.1) ou Makinson (2005, § 3.2).

⁵³Pour d'autres réserves par rapport à la possibilité de trouver des formes d'argumentation pour des contrefactuels, s'appuyant sur des considérations quelques peu différentes, voir Adams (1975, Ch IV, §8).

⁵⁴Lewis (1979b, p467).

les plus regrettables de l'hypothèse d'une distinction forte entre une partie sémantique et une partie pragmatique des contrefactuels : elle entraîne auprès des théories des contrefactuels à la fois une *dépendance excessive de la pragmatique*, débordant sur l'abus, et également un *manquement au devoir de rendre compte des aspects pragmatiques*.

Les réponses offertes à la question des semifactuels indépendants et des phénomènes de *verum ad quodlibet* qui y sont associés (§4.2.3), lesquels ont du moins *l'allure logique*, témoignent du problème. Dans chaque cas, s'ils n'ignorent pas la question, les théoriciens mettent en œuvre un dispositif pragmatique : tantôt une sélection judicieuse et particulière de relation de similitude (ou une notion de conséquence et des prémisses) relevant de la pragmatique, tantôt un effet d'indétermination ou de vague pragmatique qui enrobe la théorie sémantique précise. Si leurs théories peuvent rendre compte des particularités des semifactuels indépendants et des phénomènes liés au *verum ad quodlibet*, ou si elles peuvent du moins esquiver les éventuelles critiques sur ce point, cela n'a rien à voir avec la théorie sémantique posée : c'est bien *la pragmatique qui fait la grande partie du travail*. Or, ces théoriciens *n'offrent aucune analyse ni considération théorique relative à cette partie pragmatique cruciale*. Ils s'inscrivent à une méthodologie douteuse : ils s'appuient sur une pragmatique qu'ils ne tentent même pas de comprendre. Ils abusent d'une pragmatique qu'ils négligent par la suite.

Il se pourrait par ailleurs que le fait de négliger la question de la pragmatique soit elle-même entraînée par la distinction entre la sémantique et la pragmatique et l'idée d'un « noyau sémantique », plus précisément par la *fixation* sur ce « noyau sémantique ». Car, lorsque le regard se fixe sur une théorie au détriment de toute autre, théorie que l'on croit être *la* théorie sémantique de l'*unique* noyau sémantique, alors on croit, en confondant *logique* ou *sémantique* avec *généralité*, s'être emparé de *tous* les aspects généraux de la question. De telle sorte que l'on rejette la pragmatique comme ayant trait aux particularités dont on pourrait faire l'économie. Tout théoricien jusqu'à cette date (à l'exception discutable de Goodman) agit de telle manière ; Lewis, avec sa clarté propre, a au moins le mérite de l'affirmer explicitement : « [son analyse] est à peu près tout ce que l'on peut dire dans la généralité au sujet des contrefactuels »⁵⁵. Ce rejet n'est pas complètement justifié, comme on verra dans §§4.3.2 et 4.3.3, où seront repérés certains aspects de la pragmatique qui appellent une théorisation et qui pourraient qualifier de « généraux ».

⁵⁵ *Ibid.*

Prendre en charge la pragmatique

Le but est de construire une théorie complète des contrefactuels. Une telle théorie ne pourrait pas être *seulement* une théorie sémantique, ou *seulement* une théorie pragmatique, elle doit être une théorie du *bloc sémantique-pragmatique* en entier. Par rapport à cet objectif, le travail du théoricien qui se cloître dans un « noyau sémantique », et qui propose une théorie de ce seul noyau en ignorant la pragmatique mondaine, ne peut que décevoir. Si sa théorie a la prétention d'éclairer les contrefactuels, elle est insuffisante, dans la mesure où elle néglige leur côté pragmatique ; si sa prétention est limitée à la sémantique, elle est inintéressante, dans la mesure où elle emploie la pragmatique mystérieuse comme un *joker* pour se mettre à l'abri des intempéries de la réalité.

En recentrant la tâche de la théorie des contrefactuels pour prendre en charge aussi bien leur côté pragmatique que leur côté sémantique, on retrouve la question générale des contrefactuels posée au départ. *La question des contrefactuels*, dans sa forme générale, *est celle de leur rapport à la réalité* (§4.1.3). Il s'avère que celle-ci est *exactement* la question que l'on a négligée sous le nom de la pragmatique. La question relative au contrefactuel « si l'allumette avait été grattée, elle se serait enflammée » est exactement la question de *quelles* prémisses et *quelle* notion de conséquence, c'est-à-dire *quels faits et aspects réels*, le soutiennent (ou non). Ou encore, la question relative au contrefactuel « si l'allumette avait été grattée, alors Jacques Chirac serait le président de la France » est *exactement* la question du statut du fait réel de la présidence de Jacques Chirac relative à l'éventualité contrefactuelle de gratter l'allumette. Dans les deux cas, il est question des prémisses et de la notion de conséquence, il est question des aspects de la réalité qui entrent en jeu : il est question de l'apport pragmatique (§4.2.1).

Par ailleurs, il découle de cette réorientation de l'objectif d'une théorie des contrefactuels un certain nombre de conséquences touchant à la *forme* que devrait prendre une théorie de contrefactuels, dont deux méritent d'être soulignés. Premièrement, la question des « règles logiques » qui s'appliquent aux contrefactuels perd une large part de son intérêt ; deuxièmement, une théorie des contrefactuels ne devrait plus être globale, comme le sont les théories proposées jusqu'à présent (§4.2.1).

Souvent, les théories des contrefactuels qui ont été proposées comprennent, ou même parfois consistent en, un certain nombre de « règles logiques » ou de formes d'argumentation que les contrefactuels sont censés satisfaire⁵⁶. Or, dès

⁵⁶On a discuté de telles règles ci-dessus. Voir par exemple Lewis (1973b) ou Kraus et al. (1990). Une liste de certaines règles souvent prise pour fondamentales se trouve dans l'Annexe

que l'on vise à comprendre le problème *entier* des contrefactuels, comprenant son versant pragmatique aussi bien que logique ou sémantique, la question des règles qui s'y appliquent devient moins intéressante. Non seulement à cause de la difficulté de s'accorder sur des règles appropriées, laquelle a été constatée ci-dessus. Mais de plus, en raison de la *fonction* ou du *statut* même que l'on pourrait accorder aux règles logiques. De deux choses l'une : soit les règles logiques sont à comprendre dans un sens *normatif*, de telle sorte qu'elles n'ont aucun intérêt pour la problématique posée ici qui est, comme on l'a remarqué dans §4.1.3, *descriptive*. Soit les règles logiques ont une fonction *descriptive* ; pourtant, étant donné le décalage entre la relative simplicité des règles logiques et la complexité de ce qu'elles sont censées décrire (la pensée, le raisonnement, l'argumentation), et surtout le fait qu'elles sont parfois enfreintes *en pratique*, il faut conclure que les règles logiques sont, au mieux, des « *approximations* » ou des « *abstractions* » par rapport au (la pensée, le raisonnement, l'argumentation) réel. Pour une théorie qui travaille à un moindre niveau d'abstraction, une théorie qui se veut plus proche des faits (en l'occurrence, relatifs aux contrefactuels), les prétendues règles logiques peuvent au mieux servir de guide, et éventuellement de défi, dans la mesure où, comme l'a été mentionné dans §4.1.3, la possibilité de les reproduire par abstraction à partir de la théorie précise ne ferait qu'accroître la plausibilité de celle-ci. La théorie proposée dans le chapitre suivant admet effectivement des abstractions qui aboutissent aux règles logiques proposées traditionnellement pour les contrefactuels, comme le sera discuté dans §5.3.2 et l'Annexe A. Il n'y aura cependant *aucune raison de s'embrouiller dans les divers débats relatifs aux « bonnes » règles logiques pour les contrefactuels* ; pour celui qui cherche une théorie plus exacte, les oppositions entre les différentes abstractions ou approximations grossières sont peu intéressantes.

D'un autre côté, le fait de se charger de la pragmatique mine l'intuition selon laquelle une théorie des contrefactuels doit être globale au sens de §4.2.1, c'est-à-dire qu'elle devrait évaluer, dans tout contexte, tout contrefactuel du langage. Rien n'empêche que la théorie permet seulement l'évaluation de certains contrefactuels dans un contexte particulier – les contrefactuels qui y sont *en scène* – et que la mise en question d'autres contrefactuels est comprise comme un glissement dans un autre contexte où elle pourrait être évalué, c'est-à-dire, comme un aspect pragmatique, dont la théorie, parce qu'une théorie *complète* des contrefactuels, rend également compte. De toute façon, les théories globales des contrefactuels font à la fois *trop* et *pas assez*. *Trop* du point de vue *philosophique* : pour tout contexte seulement un sous-ensemble de tous les contrefac-

tuels du langage sont *effectivement* accordés des valeurs de vérité, de telle sorte qu'il n'y a aucune façon de mettre à l'épreuve l'évaluation des contrefactuels non appartenant à ce sous-ensemble. *Pas assez* du point de vue *technique* : tout d'abord, comme on l'a vu, les théories globales qui ont été proposées jusqu'ici rencontrent des difficultés avec le *hors sujet*, le *verum ad quodlibet*, la frontière arbitraire entre prémisses et notion de conséquence. De plus, on saurait se douter que la globalité de la théorie constitue un obstacle à la pensée de la pragmatique. Pour prendre l'exemple du changement de l'apport pragmatique (voir ci-dessous §4.3.3), la globalité exige qu'on gère un grand nombre d'énoncés (tous !) dans chaque transition d'un apport pragmatique à un autre. Le déplacement d'un monde sur l'échelle instaurée par la relation de similitude, par exemple, aura des conséquences importantes sur un grand nombre de contrefactuels et, dans la considération des changements de l'apport pragmatique (relation de similitude), il faut tenir compte de *toutes* ces conséquences et de *tous* ces contrefactuels.

En somme, la question abordée ici – la question du rapport entre contrefactuels et réalité, et donc la question des parties *pragmatique* aussi bien que *sémantique* des contrefactuels – n'est pas thématifiée par la plupart des théoriciens. Pour cette raison, alors que l'on reproduira et retrouvera certains aspects de leurs théories, il ne sera pas question de s'impliquer dans les débats spécifiques et internes, tels ceux qui opposent des partisans des différents « noyaux sémantiques » ou systèmes logiques pour les contrefactuels. Il ne s'agit pas d'entrer dans leurs locaux mais plutôt d'ouvrir la fenêtre, de remarquer qu'il y a un dehors, et un dehors très important, à savoir la pragmatique, où beaucoup reste à faire. Une fois ouverte, des aspects intéressants de la pragmatique, susceptibles d'articuler une enquête à son égard, se présentent immédiatement ; dans les deux prochains paragraphes il s'agit d'en repérer deux. Le premier aspect concerne certains rapports naturels et intuitifs entre des apports pragmatiques en vigueur dans certaines situations différentes mais reliées ; le second aspect a trait à la dynamique de la pragmatique – aux rapports entre les différents apports pragmatiques aux moments successifs. Ces aspects se posent comme des défis pour une éventuelle théorie de contrefactuels.

4.3.2 Défis externes : comparaison d'apports pragmatiques entre situations différentes

Où l'on considère deux exemples des rapports entre des contrefactuels qui apparaissent dans des contextes différents, et qui sont donc évalués selon les apports pragmatiques a priori différents.

Fait parfois négligé, il y a souvent des rapports entre différents apports pragmatiques appropriés au même contrefactuel ou aux contrefactuels apparentés dans des situations différentes. On met en évidence deux rapports par le biais de deux exemples. Le premier exemple provient de la science : on pense le plus souvent que la vérité du contrefactuel « si ce morceau de beurre avait été réchauffé à 70°C, alors il fondrait » a quelque chose à voir avec la vérité d'une loi telle que « le beurre fond à 70°C » (quels que soit les détails de ce rapport). Ainsi, on a tendance à penser que ce contrefactuel devrait être vrai dans un grand nombre de circonstances, et cela pour « la même raison ». Il suffit de comparer une situation où le beurre est à 70°C avec une situation où il ne l'est pas : on voudrait que le beurre fonde (dans la situation où il est à 70°C) pour « la même raison » qu'il *fondrait* s'il *était* à 70°C (dans la situation où il n'est pas à 70°C)⁵⁷. Ce sont des situations différentes, avec des apports pragmatiques différents, mais il y a néanmoins une communauté intuitive entre elles qui est exprimée par l'idée que les énoncés (contrefactuels ou non) relatifs au beurre sont vrais « pour la même raison ». Une tâche d'une analyse des contrefactuels qui prend en charge la pragmatique est de rendre compte de ce rapport entre apports pragmatiques.

Les contrefactuels **(E1)** et **(E2)** de §4.2.3, relatifs à la géographie des Etats-Unis si New York était dans Georgie, fournissent un autre exemple d'éventuels rapports entre apports pragmatiques relatifs aux situations différentes. Il y a des circonstances où il est approprié de dire que si New York était dans la Georgie, alors New York serait au sud **(E1)**, et d'autres circonstances où il est approprié de dire que si New York était dans Georgie, alors la Géorgie ne serait pas complètement au sud **(E2)**. Des situations différentes, des contextes différents, avec, bien entendu, des apports pragmatiques différents. Or, on connaît bien que pour un apport pragmatique qui valide l'un des contrefactuels (disons **(E1)**), l'autre (**(E2)**) est invalide. Ces apports pragmatiques sont dans un certain sens des *inverses* les uns des autres : ce qui est valide selon les uns est invalide dans les autres et inversement. Voilà encore un rapport entre les apports pragmatiques qui décident les vérités des contrefactuels qu'une théorie complète des contrefactuels devrait mettre au clair⁵⁸.

⁵⁷Dans la situation où le beurre est à 70°C, l'antécédent du contrefactuel est vrai, et on n'emploie pas normalement la construction contrefactuelle. N'empêche que, selon le sens général de contrefactuel dont il est question ici, où il exprime un « lien » entre l'antécédent et le conséquent (§4.1.1), on peut toujours considérer ce contrefactuel dans cette situation.

⁵⁸En regardant ce qu'offrent les théories discutées précédemment en matière d'analyse de ces exemples, on n'y en trouve aucune compréhension.

4.3.3 Défis externes : la dynamique

Où l'on vient comprendre que, dans la mesure où la question du rapport entre contrefactuel et réalité est effectivement celle de l'apport pragmatique en vigueur au moment approprié, et dans la mesure où l'on ne saurait répondre à cette question autrement qu'avec une explication de comment cet apport pragmatique particulier est arrivé en vigueur, la tâche la plus importante, et d'ailleurs la plus dure, d'une théorie des contrefactuels consiste à comprendre le changement de l'apport pragmatique dans le temps, c'est-à-dire sa dynamique.

Que l'on considère l'énoncé « dans telle biographie de Napoléon, Jacques Chirac est le président de France en 2004 ». À première vue, cet énoncé ressemble aux cas d'indépendance discutés dans §4.2.3, et sur la lecture des premiers deux cents pages de cette biographie, où il n'y a aucune mention de Jacques Chirac, on serait réticent de l'affirmer. Or, si dans les dernières phrases l'auteur fait allusion à l'héritage napoléonien, sans même affirmer une thèse, il serait tout à fait concevable que la question de la cinquième République et de ses présidents devienne pertinente. De la sorte, on pourrait devenir plus sûr à l'égard cet énoncé. L'apport pragmatique nécessaire aux contrefactuels a été modifié pour accorder un véritable statut à Jacques Chirac, par rapport à la biographie. Il s'agit d'un *changement d'apport pragmatique dans un mouvement dynamique*.

C'est le cas général. À un moment où deux interlocuteurs s'accordent sur (E1) (§4.2.3), le contexte fournit la prémisse selon laquelle Georgie est dans le sud. Si un troisième interlocuteur arrive et insiste sur le contrefactuel opposé (E2), et ils l'admettent, l'apport du contexte a bien été modifié : le fait que Georgie est dans le sud n'est plus prémisse, le fait que New York n'est pas dans le sud l'est à sa place. Dans la fiction, dans le discours, mais aussi bien dans la science, la gestion des croyances et même la métaphysique, le changement de contrefactuels, et donc *la dynamique des apports pragmatiques fourni par le contexte* est une question capitale. Voire même *la* question capitale.

Car, si la question des contrefactuels est celle de leur rapport à la réalité, et si cela revient effectivement à l'identification de l'apport pragmatique (les prémisses et la notion de conséquence), alors la question des contrefactuels est la question de cet apport : quel apport pragmatique (prémisses et notion de conséquence) pour ce contrefactuel dans cette circonstance ? Et on ne saurait pas *comment* répondre à cette question si ce n'est pas d'induire, sur la base des apports pragmatiques relatifs aux moments passés, celui qui est approprié au cas en question.

Si la question du rapport entre contrefactuel et réalité est difficile, c'est donc la question de la dynamique de la part pragmatique des contrefactuels qui est à l'origine de la grande part de la difficulté. *La question de la dynamique est le « noyau » de la vraie question des contrefactuels.* Ainsi, si l'on ose ambitionner une théorie des contrefactuels qui n'esquive pas leur question la plus importante, on *doit* s'efforcer d'aborder la question de la dynamique. Comme on l'a vu dans §4.3.1, les théories traditionnelles n'abordent même pas cette question, et ne la résolvent encore moins. Par contraste, le chapitre suivant propose les bases d'une théorie qui se charge de cette question, même si, comme l'exige une question d'une telle grandeur, cette théorie n'est pas achevée. Or, on ne se pousse qu'en visant à l'impossible. Ce n'est que la pensée facile qui conjugue avec la théorie facile et son questionnement facile.

4.3.4 Résumé : la question des contrefactuels

Voici un résumé des intuitions au sujet des contrefactuels repérées dans ce chapitre. Une théorie des contrefactuels devra rendre compte de ces intuitions de manière propre et satisfaisante.

1. Il y a intuitivement, pour un contrefactuel $A \Rightarrow B$, des prémisses et une notion de conséquence telles que le conséquent est conséquence de l'antécédent pris ensemble avec ces prémisses selon cette notion de conséquence (§4.2.1). Ces prémisses et cette notion de conséquence constituent la part non formelle des contrefactuels, laquelle est reconnue comme un apport pragmatique, provenant du contexte (§4.2.1).
2. La frontière entre les prémisses et la notion de conséquence n'est ni stable ni claire ni nette ; au contraire elle est « imprécise et arbitraire » (§4.2.2).
3. Les prémisses sont « indépendantes » de l'antécédent du contrefactuel, dans le sens où leurs valeurs de vérités ne sont pas « affectées » par celle de l'antécédent, même « compte tenu » des éventuelles valeurs de vérité du conséquent (§4.2.3). D'un autre côté, les contrefactuels où l'antécédent n'est pas pertinent pour le conséquent ont un statut difficile et peu clair (§4.2.3).
4. Relativement au contexte de la considération et de l'évaluation d'un contrefactuel, un grand nombre d'énoncés du langage sont *hors sujet* (§4.2.4).
5. L'apport pragmatique diffère *a priori* selon le contexte (point 1. ci-dessus) ; néanmoins, il y a de, de toute évidence, des rapports importants entre

des apports pragmatiques relatifs aux contrefactuels apparentés dans des situations différentes (§4.3.2).

6. La question de la partie « pragmatique » des contrefactuels est importante (§4.3.1). Une approche de cette question comportera, en plus des éléments du point 5., au moins une approche de la question de la dynamique de cette partie pragmatique (§4.3.3).

Dans le chapitre suivant, les bases d'une théorie des contrefactuels fidèle à ces intuitions seront proposées. De même que la discussion précédente de la question des contrefactuels, la théorie proposée s'inspire du rapprochement de la fiction, et notamment du système présenté dans le Chapitre 2.

Chapitre 5

Aborder le problème des contrefactuels

Il y a, on l'a vu dans §4.1, un rapprochement entre la *problématique générale* de la fiction et celle des contrefactuels. Il s'avère d'ailleurs que les problématiques se ressemblent de manière plus détaillée. Dans le cas de la psychologie de la fiction, il s'agissait de proposer un modèle de l'état instantané du rapport entre la fiction et la réalité qui rende compte de certaines propriétés de ce rapport à des moments particuliers *et* permette une pensée féconde de la dynamique de ce rapport (Chs. 1 et 2). La liste des défis qui se posent à une théorie des contrefactuels (§4.3.4) lui assigne une tâche semblable : de fournir un *modèle de l'apport pragmatique* qui *supporte une évaluation de contrefactuels munie de certaines propriétés* (défis 1. à 4.) *et* permette de *penser les rapports entre les apports pragmatiques différents*, notamment entre ceux qui s'enchaînent dans la *dynamique* (défis 5. et 6.).

Vu les ressemblances entre les phénomènes dont devrait rendre compte une théorie de la psychologie de la fiction et les défis pour une théorie des contrefactuels, on ne saurait pas s'étonner d'une parenté entre la théorie qui a été proposée pour celle-là (Chs. 2 et 3) et celle qui sera proposée pour celle-ci. Là où il y avait des régularités dans l'état de la conscience à un instant particulier, il fallait les modéliser avec une *structure* appropriée à cet instant ; de même, là où il y a des régularités dans les rapports entre énoncés pour un apport pragmatique individuel en vigueur à un instant particulier, il faudra poser une structure, en l'occurrence une structure « logique », qui modélise cet apport pragmatique et représente ces régularités. Là où il y avait du *hors sujet*, il fallait que cette structure soit *locale* ; de même, puisqu'il y a du *hors sujet* (§4.2.4), les structures

logiques employées seront locales. Là où il y avait des imprécisions, il fallait que la structure locale soit *vague* ou *ouverte* ; l'imprécision dans la frontière entre notion de conséquence et prémisses (§4.2.2) sera modélisée par un vague dans la structure logique locale. Enfin, là où il y avait des régularités et des phénomènes qui se rapportent à plusieurs moments distincts, il fallait une considération des *rapports entre les structures* à ces moments ; pareillement, pour rendre compte des rapports entre des apports pragmatiques différents, relatifs à des moments et à des circonstances différents, il faudra considérer les rapports entre les structures logiques locales appropriées. Plus précisément, de même qu'une exigence fondamentale sur la conception de l'état instantané de la conscience était de permettre une pensée de sa dynamique comme continue, de même le modèle proposé de l'apport pragmatique sera jugé selon sa capacité de permettre une conception continue de sa dynamique. Et, de même que l'on a su relever ce défi en pensant les structures instantanées vagues de la conscience comme des ensembles ouverts d'une topologie, il s'agira également de penser les structures logiques locales vagues comme des ensembles ouverts.

Dans la première section, une théorie des contrefactuels sera proposée qui comporte ces aspects ; ensuite, dans la deuxième section, on considérera comment elle relève les défis établis dans le chapitre précédent. Comme il a été suffisamment souligné antérieurement (§4.3.1), le questionnement auquel répond cette théorie, et notamment son attitude par rapport aux parties sémantique (ou logique) et pragmatique des contrefactuels, diffère des problématiques traditionnelles. On se douterait que la théorie offre par conséquent une perspective divergente sur de telles questions. Dans la dernière section, on considère quelques propriétés et conséquences de cette théorie, relatives à la sémantique, la pragmatique, et à ce que c'est la logique en général, et les compare aux perspectives traditionnelles.

5.1 Une théorie des contrefactuels

Le but de cette section est de proposer une conception d'apport pragmatique qui étaie une théorie des contrefactuels. L'idée sous-jacente à la proposition est simple : l'apport pragmatique approprié *est* le *contexte* en vigueur à l'instant particulier. Contexte, non pas au sens d'une quelconque *liste de paramètres « non formels »*, dont certains seulement (la relation de similitude par exemple) sont pertinents pour les contrefactuels¹. Contexte, conçu plutôt

¹C'est la conception de Lewis par exemple, selon qui la relation de similitude est une « coordonnée contextuelle » (1970, p24) ou un « composant du score conversationnel » (1979c).

comme une *structure*, comportant tous les énoncés dont il peut être question (dans ce contexte), rangés avec les rapports logiques ou linguistiques qu'ils entretiennent entre eux (en tant qu'ils figurent dans ce contexte). L'apport pragmatique pertinent pour l'évaluation des contrefactuels n'est donc pas vu comme *un paramètre tiré de ou fourni par* un contexte théoriquement plus lointain, mais plutôt comme *la totalité du contexte*, comme (« isomorphe à ») la *structure même du contexte*.

Conformément aux intuitions développées dans §4.2.1, qui mettent en rapport l'aspect « structurel » de l'apport pragmatique à l'évaluation des contrefactuels – la notion de conséquence – et les algèbres de Boole, la *structure* du contexte sera *modélisée techniquement* comme (*grosso modo*) une algèbre de Boole et *assimilée philosophiquement* à la notion de conséquence impliquée dans l'analyse des contrefactuels. Cette conception de la notion de conséquence sera développée dans §5.1.1. Quant à l'aspect « matériel » de l'apport pragmatique – les prémisses –, il sera modélisé par des éléments particuliers de l'algèbre en vigueur, comme l'a été suggéré par les considérations de §4.2.1. Une notion d'indépendance des éléments de telles algèbres, apte à rendre compte des phénomènes d'indépendance et de pertinence mis en évidence dans §4.2.3, sera proposée dans §5.1.2. Une analyse préliminaire employant ces conceptions de notion de conséquence et de prémisse, et donc s'inscrivant à la forme générale (**AGC**) (§4.2.1), sera présentée dans §5.1.3.

Cette analyse n'aura pas rendu compte de la frontière arbitraire entre notion de conséquence et prémisses (§4.2.2), ni des questions relatives aux rapports entre des apports pragmatiques différents, dont la question de leur dynamique. Pour tenir compte de la frontière arbitraire, une conception plus développée de l'apport pragmatique, que l'on appellera « contexte », sera proposée, selon laquelle il est vague ou ouvert (§§5.1.4 et 5.1.5). Ce caractère vague permettra d'entamer une théorie de la dynamique, dans la mesure où il implique une topologie et donc une notion de continuité sur les contextes.

5.1.1 La notion de conséquence et l'algèbre interprétée

Où la notion de conséquence est représentée par une structure à deux niveaux qui s'appelle algèbre interprétée. Cette structure est également conçue comme un modèle du contexte. Ainsi, puisque le contexte comporte seulement certains énoncés du langage mais non pas tous, comme le témoigne le phénomène de hors sujet, il faut que la structure qui le modélise explicite les énoncés du langage qui sont en jeu. L'algèbre interprétée fait exactement ceci : un niveau

correspond aux énoncés localement en jeu, l'autre correspond à la structure logique sur ces énoncés dans ce contexte.

La notion de conséquence qui figure dans la forme générale d'analyse des contrefactuels (**AGC**) constitue l'aspect « *structurel* » dans l'analyse de contrefactuels (§4.2.1). Comme l'a été discuté dans §4.2.1, il y a deux manières techniquement équivalentes mais heuristiquement distinctes de concevoir cet aspect « structurel ». Selon la manière extensionnelle, elle a affaire à tous les « mondes » ou « états », alors que, selon la manière algébrique, elle a affaire à l'ordre de l'algèbre de Boole. La théorie des contrefactuels qui sera proposée ici est de forme générale (**AGC**) ; en outre, la notion de conséquence qui figure comme partie de l'apport pragmatique à cette analyse sera *apparentée* aux structures traditionnelles employées pour modéliser, représenter ou (selon certains) « être » la conséquence, à savoir un ensemble de « mondes » ou « états », ou, de manière équivalente, une algèbre de Boole. La seule différence avec les structures employées traditionnellement est appelée par l'observation du phénomène de *hors sujet*.

Dans §4.2.4, on a vu que non seulement il y a des énoncés du langage qui ne sont pas *en jeu* à des moments particuliers, mais que ce phénomène pourrait être compris comme un genre d'*absence* de ces énoncés par rapport à l'algèbre de Boole appropriée. Pour exprimer cette propriété en termes extensionnels, il s'agit des énoncés qui n'ont aucune valeur de vérité dans aucun « monde » (§4.2.1) ; comme remarqué dans §4.2.1, il est plus correct de parler de « petits mondes » que de « mondes possibles » au sens propre du terme. La structure de base du contexte qui constituera l'apport pragmatique dans l'analyse proposée des contrefactuels sera celle d'une notion de conséquence dont il y a des énoncés *hors sujet*, c'est-à-dire une algèbre de Boole comportant seulement certains énoncés du langage, ou de manière équivalent, un ensemble de « petits mondes » où seulement certains énoncés sont accordées des valeurs de vérité.

Cette prise en charge du phénomène de *hors sujet* exige une attention spéciale au langage et aux énoncés. Les énoncés qui sont *en jeu* changent de l'un contexte à l'autre, de telle sorte qu'il devient important d'*explicitement* les énoncés qui sont impliqués ou présents dans un contexte. Il ne suffit pas de supposer un langage donné fixe, comme on le fait normalement lors des considérations des questions logiques ou linguistiques. C'est dans l'explicitation des énoncés *en jeu* au moment particulier, dans le contexte particulier, que le modèle du contexte, de la notion de conséquence, et de l'apport pragmatique à l'évaluation des contrefactuels diverge des modèles ou structures employés communément.

On appelle l'ensemble d'énoncés qui sont *en jeu* dans un contexte particulier,

c'est-à-dire le fragment du langage qui est localement effectif, le *langage local*. On suppose dans ce qui suit qu'il est clos sous les opérations booléennes : par exemple, si deux énoncés A et B sont *en jeu*, alors l'énoncé A et B est *en jeu*. Plus formellement, le langage local est l'ensemble d'énoncés engendrés à partir d'un ensemble d'énoncés pris (localement) pour atomiques, I , à l'aide d'opérations booléennes (*et*, *ou*, *non* et ainsi de suite). En supposant les tautologies ordinaires relatives aux opérations booléennes, on considère ce langage comme l'*algèbre de Boole* engendrée par l'ensemble d'énoncés atomiques I . On appelle cette algèbre de Boole B_I .²

La « *logique* » (locale) sur ces énoncés, la notion de conséquence sur eux consistera en une autre structure, en l'occurrence une autre algèbre de Boole appelée B , et une fonction q munie de certaines propriétés (elle est un homomorphisme surjectif d'algèbres de Boole), qui assigne à chaque énoncé un élément de cette structure. On obtient donc une structure à deux niveaux, avec deux algèbres de Boole et une fonction appropriée entre eux, que l'on appelle *algèbre interprétée* (ou parfois juste algèbre). (Comme l'a été noté dans §4.1.2, les détails techniques seront reportés à la fin des paragraphes ; dans le cas présent, voir la Définition 5.1.3 ci-dessous.) L'algèbre interprétée est la structure de base dans la définition du contexte qui constituera l'apport pragmatique de la théorie des contrefactuels proposée. Elle pourrait se comprendre de manière extensionnelle ou algébrique.

Dans une perspective extensionnelle, on pourrait prendre un ensemble de « (petits) mondes » et une fonction d'interprétation qui assigne à chaque énoncé du langage local une valeur de vérité dans chaque « monde ». Le fait que l'interprétation se comporte de manière appropriée relative aux connecteurs booléens implique que l'interprétation est une fonction bien définie et munie des propriétés appropriées, prenant les éléments de l'algèbre B_I dans l'ensemble de petits mondes. Par contraste, dans une perspective algébrique, on représente la structure logique sur les énoncés du langage local par l'algèbre de Boole B ; la fonction (q) assigne à chaque énoncé un élément de cette algèbre qui lui « correspond ». Les énoncés (localement) équivalents correspondent aux mêmes éléments de cette algèbre ; si un énoncé implique un autre (selon la logique du contexte), il correspond à un élément de l'algèbre qui est au-dessous de l'élément qui correspond à l'autre énoncé dans l'algèbre.

Ces deux manières de se figurer la structure logique sont équivalentes : les

²La décision de prendre une algèbre de Boole, plutôt qu'une algèbre avec la même signature mais sans la structure booléenne (ce que Blackburn et al. (2001, §5.1) appellent « l'algèbre de formules »), tient à des facilités techniques qu'elle impliquera dans la suite. Le choix inverse aboutira aux mêmes résultats, mais exigera des définitions plus subtiles et complexes.

éléments de l'algèbre B correspondent aux ensembles de « petits mondes » dans la conception extensionnelle, comme on l'a vu dans §4.2.1³. En ce sens, l'algèbre B est l'algèbre de « *propositions* », alors que B_I est l'algèbre d'« *énoncés* ». B est la « sémantique » locale pour ainsi dire, alors que B_I est le « syntaxe » locale. On appelle l'algèbre d'« *énoncés* » (B_I) *algèbre d'interprétation* et l'algèbre de propositions (B) *algèbre de base*.

Le fait que l'algèbre interprétée modélisant le contexte comporte cette double structure, qu'il comporte le « syntaxe » aussi bien que la « sémantique » pour ainsi dire, a des conséquences importantes. Plus notamment, puisque le seul « langage » en jeu est le langage local, consistant seulement en les énoncés de l'algèbre d'interprétation, *aucun langage pré-donné* n'est supposé. Une algèbre interprétée a son propre langage « *interne* », si l'on peut dire. Par conséquent, les structures employées dans l'analyse des contrefactuels qui sera proposée ne sont pas interprétées relatives à un langage *extérieur à elles*, mais comportent leurs langages propres et leurs interprétations *en elles*. Il s'ensuit que le cadre théorique proposé ici représente une nette généralisation des cadres théoriques employés habituellement : comme cas spécial, on peut supposer un langage fixe pré-donné, et identifier les énoncés figurant dans les algèbres interprétées avec des énoncés de ce langage fixe.

Certains points terminologiques permettront d'alléger la discussion suivante. Pour parler des *éléments* d'une algèbre interprétée, on parle de ses *énoncés*⁴. La notion de *conséquence* sur ces énoncés est celle qui subsiste sur les propositions correspondant aux énoncés : on dit que B est conséquence de A si la proposition correspondant à A implique celle correspondant à B (Définition 5.1.5). On écrit $A \leq B$ pour dire que A est conséquence de B en ce sens, et $A \equiv B$ pour l'équivalence. Finalement, une *valuation* des énoncés est une fonction qui prend chaque énoncé à une valeur de vérité (*Vrai* ou *Faux*), en respectant la structure logique sur eux (Définition 5.1.6). Comme l'a été discuté dans §4.2.1, il s'agit de la désignation d'un « petit monde » actuel dans la conception extensionnelle, et d'une fonction appropriée sur l'algèbre B dans la conception algébrique.

La figure 5.1 et le tableau 5.1 résument la situation, dans les deux conceptions.

Pour conclure ce paragraphe, il convient de rejeter tout de suite une objection potentielle à ce choix d'apport pragmatique, laquelle tient au fait que, relatif à une algèbre qui manque certains énoncés (d'un langage fixe pré-donné),

³ B est donc supposé atomique. En pratique, puisque le langage local est fini, B est elle aussi finie, alors que cette propriété ne jouera pas de rôle dans la suite.

⁴Ceci n'est pas ambigu puisqu'une unique proposition correspond à chaque énoncé

FIG. 5.1 – Algèbre interprétée

$$\text{Algèbre interprétée} \left\{ \begin{array}{c} B_I \\ \downarrow q \\ B \end{array} \right.$$

TAB. 5.1 – Algèbre interprétée dans les deux conceptions

	Conception algébrique	Conception extensionnelle
B_I	Énoncés du langage local	
B	Algèbre de propositions	(Algèbres d'ensembles de) « petits mondes »
q	Fonction d'énoncés aux propositions	Interprétation des énoncés dans « petits mondes »
Élément	Énoncé et proposition correspondant	Énoncé et ensemble de « petits mondes » correspondant
Conséquence	Conséquence sur les propositions	Implication matérielle vraie dans tous les « petits mondes »
Valuation	Valuation des énoncés	Désignation d'un « petit monde actuel »
—	—	Monde possible

il y a des contrefactuels qui ne peuvent pas être évalués, à savoir ceux dont l'antécédent ou le conséquent ne figurent pas dans l'algèbre. Pour employer le terme introduit dans §4.2.1, une théorie des contrefactuels qui emploie un ap-

port pragmatique comportant ce type de notion de conséquence ne pourrait pas être *globale*. Pourtant, cette objection a été déjà désamorcée dans §4.3.1, où il a été constaté que, une fois que l'on accepte de prendre en charge le côté pragmatique des contrefactuels, il n'est plus besoin qu'une théorie des contrefactuels soit globale; d'ailleurs, la globalité constitue un net inconvénient à une telle théorie. La théorie de contrefactuels qui sera développée ici ne sera donc pas globale, dans la mesure où seulement certains contrefactuels seront évalués dans un contexte particulier; elle sera *locale*.

Voilà quelques définitions techniques qui permettent une compréhension formelle de la discussion précédente.

Définition 5.1.1 (Algèbre de Boole). Une *algèbre de Boole* est un ensemble d'éléments B équipé d'un ordre partiel \leq , deux éléments spéciaux \top, \perp , deux opérateurs binaires $\wedge, \vee : B \times B \rightarrow B$, et un opérateur unaire $\neg : B \rightarrow B$ où :

- \leq est un treillis avec (pour tout x, y) $\sup(x, y) = x \vee y$, $\inf(x, y) = x \wedge y$;
- \wedge et \vee sont distributifs l'un sur l'autre ($x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ pour tout x, y);
- pour tout x , $\perp \leq x \leq \top$;
- pour tout x , $x \wedge \neg x = \perp$, $x \vee \neg x = \top$.

On emploie \leftrightarrow pour le connecteur $A \leftrightarrow B = (A \wedge B) \vee (\neg A \wedge \neg B)$.

L'algèbre d'interprétation, laquelle modélise le langage local, est une algèbre *engendrée* par un ensemble d'éléments I . On appelle de telles algèbres des algèbres libres, et I une *base* d'une telle algèbre.

Définition 5.1.2 (Algèbre libre). L'*algèbre de Boole libre* engendrée par un ensemble I est l'algèbre de Boole B_I dont (1) les éléments sont \perp, \top et les éléments formées à partir des éléments de I avec les opérateurs \wedge, \vee, \neg , et (2) les seules relations \leq entre éléments sont celles qui sont impliquées par les axiomes d'algèbres de Boole (Définition 5.1.1).⁵

On pourrait donc définir une algèbre interprétée.

Définition 5.1.3 (Algèbre interprétée). Une *algèbre interprétée* \mathbf{B} est la donnée de :

- une algèbre libre B_I (la dite *algèbre d'interprétation*)
- une algèbre B (la dite *algèbre de base*)
- un homomorphisme surjective $q : B_I \rightarrow B$

Parfois, on note l'algèbre interprétée \mathbf{B} par le triplet (B_I, B, q) .

⁵Plus techniquement, B_I est le quotient de l'algèbre de termes (ou de formules) (Blackburn et al., 2001, §§5.1 et B) sur I par les équivalences qui sont théorèmes de la théorie des algèbres de Boole.

Remarque 5.1.1. Grâce à des résultats algébriques bien connus, puisque q est une homomorphisme surjective, elle est également l'application de quotient d'un quotient d'algèbre de Boole, définie de manière suivante.

Définition 5.1.4 (Quotient). Pour un élément x d'une algèbre de Boole B , le quotient de B par (l'idéal engendré par) x , $B/(x)$, est l'algèbre obtenue de B en soustrayant tous les éléments $y \leq x^6$. On appelle *application de quotient* l'homomorphisme surjectif $q_x : B \rightarrow B/(x)$ associant à chaque élément à son image par cette soustraction (ou sa classe d'équivalence).

De manière pareille, B/\mathcal{G} pour une collection d'éléments \mathcal{G} est le quotient par tous les éléments, avec l'application de quotient $q_{\mathcal{G}}$.

En l'occurrence, $q : B_I \rightarrow B$ dans la définition de l'algèbre interprétée est l'application de quotient pour le quotient par $\{A \in B_I \mid q(A) = \perp\}$. On appelle cet ensemble \mathcal{F} , et on parle parfois de l'algèbre de base comme le quotient de l'algèbre interprétée par \mathcal{F} .

On parle des éléments d'une algèbre interprétée, et de la relation de conséquence entre ces éléments, selon la définition suivante.

Définition 5.1.5 (Élément et conséquence pour algèbres interprétées). Un élément A de l'algèbre interprétée (B_I, B, q) est un élément $A \in B_I$.

Un élément B de (B_I, B, q) est conséquence de l'élément A , écrit $A \leq B$ ssi $q(A) \leq q(B)$. Également, A et B sont équivalents, $A \equiv B$, ssi $q(A) = q(B)$.

Cette définition de valuation est équivalente de celle utilisée ci-dessus.

Définition 5.1.6 (Valuation). Une *valuation* sur une algèbre interprétée (B_I, B, q) est un homomorphisme $v : B_I \rightarrow \mathbf{2}$, où $\mathbf{2}$ est l'algèbre de Boole à deux éléments (appelés V et F) et où $v = v' \circ q$, pour un homomorphisme $v' : B \rightarrow \mathbf{2}$.

Une valuation d'un ensemble E d'énoncés de (B_I, B, q) est une valuation de la sous-algèbre engendrée par E .

5.1.2 La prémisse et la co-indépendance

Où une notion de co-indépendance est définie sur les ensembles d'éléments des algèbres interprétées, s'inspirant de la notion algébrique d'indépendance. Les prémisses dans l'analyse de contrefactuels seront des ensembles des éléments qui sont co-indépendants dans ce sens rigoureux. L'exigence que les prémisses et l'antécédent soient co-indépendants compte tenu du conséquent correspond à l'exigence que le conséquent soit équivalent, dans

⁶A strictement parler, c'est l'algèbre de classes d'équivalence de la relation $y \sim z$, où $y \sim z$ ssi $y \triangle z \leq x$, où \triangle est la différence symétrique.

l'algèbre interprétée, à une combinaison logique des prémisses et de l'antécédent. L'apport pragmatique dans l'analyse préliminaire des contrefactuels, que l'on appelle cadre, consistera donc en une algèbre interprétée et un ensemble co-indépendant de ses énoncés.

Il faut maintenant venir à l'autre aspect de l'apport pragmatique requis, selon la forme générale **(AGC)**, pour évaluer les contrefactuels, à savoir les prémisses, ou autrement dit l'aspect « matériel » (§4.2.1). Elles seront comprises ici comme *certaines énoncés* de l'algèbre interprétée – c'est-à-dire certains énoncés en jeu dans le contexte présent. Or, ce n'est pas vrai que n'importe quel ensemble d'énoncés ferait l'affaire : en l'occurrence, on a remarqué dans §4.2.3 deux relations entre les prémisses, l'antécédent et le conséquent — les prémisses d'un contrefactuel constituent avec l'antécédent un *ensemble co-indépendant* — cet ensemble est co-indépendant *compte tenu* du conséquent du contrefactuel. Il faut donner à ces notions un sens dans l'algèbre interprétée : on le fait premièrement pour la notion de « co-indépendance », et ensuite pour « co-indépendance compte tenu de ».

Pour aborder la notion de co-indépendance, on dispose déjà d'une notion issue de la mathématique : si un ensemble d'éléments E est tel que la collection des relations mathématiques (non triviales) qu'ils entretiennent les uns avec les autres est vide, alors ces éléments sont *indépendants ensemble* relatifs les uns aux autres, dans le sens où la valeur que l'on assigne à l'un n'a aucune conséquence sur les valeurs que l'on peut assigner aux autres. La définition d'indépendance sur les algèbres interprétées s'inspire de cette définition (pour la version technique, voir Définition 5.1.7 ci-dessous) :

Un ensemble d'énoncés E d'une algèbre interprétée \mathbf{B} est un ensemble co-indépendant si et seulement si il n'y a aucune relation non triviale dans l'algèbre entre les énoncés de E .

Il découle immédiatement de cette définition que, pour un énoncé A d'un ensemble co-indépendant E , aucune valuation des autres énoncés de E (c'est-à-dire, aucune assignation des valeurs de vérité V ou F aux autres énoncés) n'entraînerait des contraintes sur la valuation de A : car, si c'était le cas, alors il y aurait des relations non triviales dans l'algèbre reliant seulement les énoncés de E . Cette indépendance des valuations des énoncés d'un ensemble co-indépendant les uns par rapport aux autres est exigée par l'analyse de §4.2.3, où l'on a remarqué que les valeurs de vérité des prémisses ne devraient dépendre d'aucune manière de celle de l'antécédent (ni, d'ailleurs, d'autre prémisses).

Cette conception de co-indépendance appelle trois remarques. D'abord, il peut y avoir plusieurs ensembles co-indépendants dans une même algèbre, comme le montre l'algèbre engendrée par trois énoncés A , B et C avec $C \equiv A \vee B \equiv (A \leftrightarrow \neg B) \vee A$ (la seconde équivalence est un fait algébrique). $\{A, B\}$ et $\{A, A \leftrightarrow \neg B\}$ sont deux ensembles co-indépendants *différents* de cette algèbre⁷. En vérité, il est un fait mathématique qu'il peut y avoir plusieurs ensembles co-indépendants dans une algèbre. En outre, les différents ensembles co-indépendants sont généralement *incompatibles* entre eux, en ce sens que leur union n'en est pas un ensemble co-indépendant. Dans l'exemple, l'union des ensembles co-indépendants, $\{A, B, A \leftrightarrow \neg B\}$, n'est pas un ensemble co-indépendant (une valuation particulière de A et B implique une valuation particulière de $A \leftrightarrow \neg B$).

Ensuite, cette notion de co-indépendance ne concorde pas seulement avec la notion algébrique, dont il tire une inspiration, mais également avec la notion d'indépendance ou de « non pertinence » probabiliste privilégiée dans certains écrits philosophiques, selon laquelle, dans sa forme la plus simple, A est « non pertinent » à C si $P(C/A) = P(C)$ où $P(A)$ est la probabilité de A ⁸. Pour voir ce rapport, il suffit de mobiliser l'idée wittgensteinienne de « probabilité logique »⁹, selon laquelle la probabilité d'un énoncé est la proportion d'atomes de l'algèbre de Boole qui impliquent l'énoncé¹⁰. Il n'est pas difficile de vérifier que la notion d'indépendance algébrique en œuvre dans ce travail *implique* la notion traditionnelle de non pertinence (formulée avec la « probabilité logique » wittgensteinienne) : si A et C sont indépendants au sens algébrique, alors ils sont non pertinents selon la définition probabiliste.¹¹

Enfin, cette conception de co-indépendance permet une notion claire et simple de « co-indépendance compte tenu de ». Il y a, pour un ensemble co-indépendant E et un énoncé X de l'algèbre (non pas équivalent au vrai ou au faux), quatre statuts possibles de la relation entre E et X :

1. X appartient à E ;

⁷Remarquons que cet exemple illustre de surcroît qu'il peut y avoir deux ensembles co-indépendants différents qui sont les *bases différentes* de la même algèbre libre, ce qui est d'ailleurs un fait mathématique bien connu (pour la notion de la base d'une algèbre libre, voir la Définition 5.1.2, §5.1.1).

⁸Voir par exemple Carnap (1950).

⁹Wittgenstein (1922, 5.15 *sq.*). Voir aussi Carnap (1950).

¹⁰Rappelons que les algèbres de base ici sont atomiques (note 3).

¹¹Pour autant qu'elles soient correctes, d'autres notions de non pertinence probabilistes, supposément plus fortes, sont impliquées par l'indépendance algébrique. Pour un exemple d'une telle notion, voir Gärdenfors (1978) (malheureusement sa notion est triviale, dans la mesure où elle ne permet pas trois énoncés indépendants chacun de probabilité $\frac{1}{2}$: pour A , B , et C de tels énoncés, considérer $D = (A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge C) \vee (\neg A \wedge B \wedge C) \vee (A \wedge \neg B \wedge \neg C)$).

2. X n'appartient pas à E mais il est équivalent à une combinaison logique des énoncés de E ;
3. X n'appartient pas à E , n'est pas équivalent à une combinaison logique de ses énoncés mais $E \cup \{X\}$ (le résultat d'ajouter X à E) n'est pas co-indépendant ;
4. même que 3. sauf que $E \cup \{X\}$ est co-indépendant.

L'algèbre \mathbf{B} engendrée par A , B et G , avec $A \wedge G \leq B$, qui permet $\{A, G\}$ comme ensemble co-indépendant (le E), suffit pour illustrer ces différences. On trouve un exemple de cas 1. en prenant $X = A$, de cas 2. en prenant $X = A \wedge G$, et de cas 3. en prenant $X = B$, car $\{A, G, B\}$ n'est pas co-indépendant¹². Pour la question de la relation entre le conséquent d'un contrefactuel (X) et l'ensemble constitué des prémisses et de l'antécédent (E), seulement les cas 2. et 3. sont intéressants. Le phénomène de l'indépendance « compte tenu de » que l'on a mis en évidence dans §4.2.3 tient à « l'effet » de la « mise en jeu » du conséquent sur la co-indépendance de l'ensemble consistant en des prémisses et l'antécédent. Il y a un tel « effet » dans le cas 3. mais non pas dans le cas 2.

On peut dire que les énoncés d'un ensemble S entrent dans une relation non triviale *modulo* Y lorsqu'il y a une relation non triviale reliant les énoncés de $S \cup \{Y\}$ qui ne peut pas être exprimée en termes des énoncés de S seul. Dans le cas 2., les énoncés de E n'entrent pas dans une relation non triviale *modulo* X : puisque X est une combinaison logique des énoncés de E , tout ce qui peut être exprimé en employant X peut être exprimé déjà en termes des énoncés de E . X n'ajoute pas de relations nouvelles, si l'on peut dire. En revanche, dans le cas 3., les énoncés de E entrent dans des relations non triviales *modulo* X : puisque X n'est pas une combinaison logique des énoncés de E , il y a des relations non triviales entre X et les énoncés E ($E \cup \{X\}$ n'est pas co-indépendant) qui ne sont pas exprimables en termes des énoncés de E seuls. Autrement dit, *modulo* (une valuation de) X , les valeurs assignées à certains éléments de E peuvent avoir des conséquences sur les valeurs qui peuvent être accordées à d'autres éléments de E . En ce sens, E n'est pas co-indépendant « compte tenu de X ».

Voilà l'explication algébrique de la discussion pré-théorique de §4.2.3, qui est arrivée à la conclusion que l'ensemble co-indépendant consistant en des prémisses et de l'antécédent est co-indépendant « compte tenu du conséquent ». En termes algébriques, il faut être dans le cas 2. plutôt que 3. : le conséquent devrait être *équivalent* (dans l'algèbre) à une combinaison logique des prémisses et de l'antécédent.¹³

¹²Pour le cas 4., on peut prendre par exemple, $E = \{A\}$, et $X = G$.

¹³On remarque par ailleurs que, dans le cas 2., puisque X est une combinaison logique des

Dans le paragraphe précédent, on a identifié un aspect de l'apport pragmatique pour une analyse des contrefactuels comme une algèbre, fournissant la notion de conséquence. L'autre aspect sera un ensemble co-indépendant, qui fournit les prémisses de l'analyse. La donnée d'une algèbre interprétée (modélisant la notion de conséquence) et un ensemble co-indépendant d'énoncés de cette algèbre (modélisant les prémisses) s'appelle un *cadre* (Définition 5.1.8).

Techniquement, on définit un ensemble co-indépendant d'une algèbre interprétée en s'appuyant sur la notion d'algèbre de Boole libre (Définition 5.1.2).

Définition 5.1.7 (Co-indépendance). Pour un ensemble d'éléments $E = \{A_i\}$ d'une algèbre interprétée (B_I, B, q) , soit B_E l'algèbre libre engendrée par $\{q(A_i) \mid A_i \in E\}$ (les images des éléments de E dans B). E est *co-indépendant* si et seulement si $B_E \subseteq B$.

Remarque 5.1.2. Puisque, pour toute paire d'ensembles $E = \{A_i\}$ et $E' = \{A'_i\}$ où, pour tout i , $A_i = A'_i$ ou $A_i = \neg A'_i$, on a $B_E = B_{E'}$, la co-indépendance d'un ensemble est préservée sous l'opération de prendre les négations de n'importe quel sous-ensemble de ses éléments. La notion d'ensemble co-indépendant employée dans cette thèse est effectivement utilisée à la *négation de ses éléments près* : on ne fera pas de distinction entre les ensembles $\{A, B, C\}$ et $\{\neg A, B, \neg C\}$ par exemple. En pratique, on supposera le plus souvent que c'est l'antécédent du contrefactuel et les prémisses (évaluées comme vraies) qui figurent dans l'ensemble co-indépendant (plutôt que leurs négations).

L'exigence que le conséquent C soit équivalent à une combinaison logique des prémisses P_i et de l'antécédent A est, techniquement, l'exigence que $q(C) \in B_{\{q(P_i)\} \cup \{q(A)\}}$. On raccourcit cette notation lourde en écrivant $C \in B_{\{P_i\} \cup \{A\}}$.

Définition 5.1.8 (Cadre). Un *cadre* est une paire (\mathbf{B}, E) où \mathbf{B} est une algèbre interprétée et E est un ensemble co-indépendant d'éléments de \mathbf{B} .

5.1.3 Une analyse préliminaire des contrefactuels

*Où une analyse préliminaire des contrefactuels, prenant comme apport pragmatique un cadre, sera proposée. Cette analyse est de la forme générale (**AGC**) : la notion de conséquence est celle livrée*

énoncés de E , sa valuation est *déterminée* par les valuations des énoncés de E , alors que, dans le cas 3., où X n'est pas une combinaison logique, il y a des valuations des énoncés de E telles que la valuation de X n'est pas déterminée. Autrement dit, dans cas 2., X est (totalement) « *dépendant* » de E , alors que dans le cas 3., il ne l'est pas ; ce qui rejoint les remarques de §4.2.3 relatives à la « *dépendance* » du conséquent par rapport à l'ensemble consistant en l'antécédent et des prémisses.

par l'algèbre interprétée, les prémisses sont les énoncés de l'ensemble co-indépendant.

On est maintenant en position de formuler une analyse préliminaire de contrefactuels, relative à la donnée d'un *cadre*, c'est-à-dire d'une paire consistant en une algèbre interprétée et un ensemble co-indépendant d'énoncés de cette algèbre. La *notion de conséquence* employée dans l'analyse consistera en les relations structurelles de l'algèbre interprétée; pour qu'il y ait une relation quelconque entre eux (selon la notion de conséquence), l'algèbre doit donc comporter l'antécédent et le conséquent du contrefactuel comme énoncés. Les *prémisses* requises par l'analyse seront les énoncés de l'ensemble co-indépendant autres que l'antécédent; de fait, l'antécédent doit appartenir à l'ensemble co-indépendant. De plus, comme on a montré dans §5.1.2, pour assurer que cet ensemble soit co-indépendant *compte tenu du conséquent*, il faut que le conséquent soit équivalent à une combinaison logique des énoncés de cet ensemble. Pour assurer la pertinence de l'antécédent au conséquent, on suppose que cette combinaison logique figure nécessairement l'antécédent (on revient en détail sur cette supposition dans §5.2.1). L'analyse locale préliminaire ((ALP)) des contrefactuels est comme suit.

(ALP). Soit la donnée d'un cadre (\mathbf{B}, E) avec une valuation v des éléments de E , où 1. l'algèbre interprétée \mathbf{B} contient les énoncés A et B , 2. l'ensemble co-indépendant $E = \{P_1, \dots, P_n, A\}$ avec P_1, \dots, P_n tous évalués comme vrais selon v ¹⁴, et 3. B est équivalent à une combinaison logique des énoncés de E , figurant nécessairement A . Alors

$A \Rightarrow B$ est vrai (relatif à (\mathbf{B}, E) et v) si B est conséquence (dans \mathbf{B}) de A prise en conjonction avec les P_i .

Formellement, en employant la notation \vdash introduite dans §4.2.1 pour expliciter le rôle de l'apport pragmatique :

(ALP). Pour un cadre (\mathbf{B}, E) avec une valuation v des éléments de E , où

1. $A, B \in \mathbf{B}$;
2. $A \in E$ et $\forall P_i \in E \setminus \{A\}, v(P_i) = V$;
3. $B \in B_E$, mais $B \notin B_{E \setminus \{A\}}$ ¹⁵;

¹⁴Puisque, comme l'a été constaté dans la Remarque 5.1.2, la notion d'ensemble co-indépendant est définie à la négation de ses énoncés près, la condition que les énoncés de E sont évalués comme vrais n'exige que l'on choisisse celui entre P_i et $\neg P_i$ qui est évalué comme vrai.

¹⁵Voir les remarques terminologiques de §5.1.2.

$$\vdash_{(\mathbf{B}, E), v} A \Rightarrow B \text{ si } A \wedge P_1 \wedge \cdots \wedge P_n \leq B$$

Deux exemples serviront à mettre l'idée au clair.

Exemple 5.1.1. En reprenant l'exemple de l'allumette, on a le contrefactuel $A \Rightarrow B$, où

- A = l'allumette est grattée
- B = elle s'enflamme
- G = elle est sèche
- Et on a que $A \wedge G \leq B$.

L'algèbre \mathbf{B}_1 engendrée par ces trois énoncés avec $A \wedge G \leq B$ n'a pas d'ensembles co-indépendants intéressants qui satisfont aux conditions de **(ALP)**. L'analyse exige donc le passage dans une algèbre qui permet un ensemble co-indépendant convenable, dont un exemple simple est l'algèbre \mathbf{B}_2 comportant un énoncé générateur supplémentaire, R , avec la formule $B \equiv (A \wedge G) \vee R$, laquelle permet l'ensemble co-indépendant $\{A, G, R\}$. Relatif à ce cadre (c'est-à-dire cette algèbre et cet ensemble co-indépendant), le contrefactuel est validé, selon analyse **(ALP)**, relatif à ce cadre¹⁶.

Il y aura l'occasion de s'attarder sur les contraintes imposées par l'analyse sur le cadre dans §5.2.1 ; quelques remarques au sujet de ce R (pour « reste ») suffiront pour le moment. Intuitivement, il est la disjonction de toutes les autres manières que l'allumette s'enflammerait – les incendies, les coups de foudre et ainsi de suite – et comme tel, il figure déjà dans d'autres analyses des contrefactuels (avec leurs propres conceptions d'apports pragmatiques). Dans l'analyse en termes de mondes possibles par exemple, $R \setminus (A \wedge G)$ est l'ensemble de mondes où l'allumette s'enflamme (B) mais soit elle n'est pas grattée ($\neg A$), soit elle n'est pas sèche ($\neg G$) ; cet ensemble n'est pas vide mais « loin » du monde actuel selon la relation de similitude. La particularité de l'analyse **(ALP)** tient au fait que, en raison de la notion algébrique d'indépendance et l'exigence que le conséquent soit une combinaison logique des énoncés de l'ensemble co-indépendant, ce « reste » est explicité comme énoncé.¹⁷

¹⁶Ici, comme dans la plupart des exemples discutés dans ce travail, il y a une valuation « évidente ». Dans ces cas, la valuation « évidente » sera supposée.

¹⁷Il est important de remarquer qu'il existe un cadre sans reste qui évalue le contrefactuel correctement (et qui ne permet pas le contrefactuel $A \Rightarrow \neg G$), à savoir le cadre consistant en l'algèbre engendrée par A , B et G avec $B \equiv A \wedge G$, et en l'ensemble co-indépendant $\{A, G\}$. Dans la structure logique (locale) représentée par ce cadre, l'allumette s'enflamme *si et seulement si* elle est grattée et elle est sèche. La simplicité de ce genre de cadre, quoiqu'elle nuise à son exactitude en tant que modèle des cadres en vigueur en pratique, permet d'alléger les discussions techniques, si bien que ce genre de cadre sera utilisé à des moments où l'emploi d'un cadre avec un reste aboutirait aux mêmes conclusions mais compliquerait l'exposition.

Exemple 5.1.2. Un autre exemple est le contrefactuel $\neg A \Rightarrow C$ (« si Oswald n'avait pas tué Kennedy, Kennedy aurait été tué »), où

- A = Oswald a tué Kennedy
- B = Quelqu'un d'autre a tué Kennedy
- C = Kennedy a été tué
- Et on a que $C = A \vee B$.

L'algèbre \mathbf{B} engendrée par A , B , et C avec $C \equiv A \vee B$ a été discutée dans §5.1.2, où l'on a remarqué qu'elle permet les deux ensembles co-indépendants $\{A, B\}$ et $\{A, A \leftrightarrow \neg B\}$. Si l'on emploie le cadre $(\mathbf{B}, \{A, B\})$, le contrefactuel $\neg A \Rightarrow C$ est évalué comme faux ; de l'autre côté, si l'on emploie le cadre $(\mathbf{B}, \{A, A \leftrightarrow \neg B\})$, le contrefactuel est évalué comme vrai¹⁸. Cet exemple montre que différents ensembles co-indépendants d'une même algèbre peuvent résulter dans des analyses différentes.

La notion de cadre, et l'analyse (**ALP**) qui l'emploie, réussissent, comme on le verra dans §5.2.1, à rendre compte d'un certain nombre de propriétés des contrefactuels qui ont été repérées dans le chapitre précédent. Pourtant, il y a toujours des propriétés importantes qui leur échappent. En particulier, le caractère imprécis et arbitraire de la frontière entre notion de conséquence et prémisses (§4.2.2, §4.3.4 défi 2.) manque : chaque cadre particulier correspond à une division précise entre la notion de conséquence (en l'occurrence l'algèbre interprétée) et les prémisses (en l'occurrence l'ensemble co-indépendant). Pour rendre compte de ce phénomène, il paraît nécessaire de remplacer des cadres précis par des structures *vagues*.

Or, si l'imprécision dans la frontière entre notion de conséquence et prémisses indique un vague dans l'apport pragmatique pertinent aux contrefactuels, il ne s'ensuit aucunement que cet apport pragmatique est sans structure aucune. Au contraire, s'il n'y avait rien qui pourrait être dit de l'apport pragmatique vague, il serait incapable de soutenir n'importe quelle évaluation de n'importe quel contrefactuel. Dans le paragraphe suivant, il s'agit de repérer une *certaine structure* dans le vague ou l'imprécision relative à la frontière entre la notion de conséquence et les prémisses : pour dire vite, la frontière est vague, mais seulement selon *une certaine dimension*. La notion d'apport pragmatique qui figurera dans la version finale de l'analyse des contrefactuels, que l'on appellera

¹⁸Cet exemple est voisin de celui évoqué fameusement par Adams (1970) pour soutenir une distinction (logique) entre les conditionnels indicatifs (« si Oswald n'a pas tué Kennedy, alors quelqu'un d'autre l'a fait ») et les conditionnels subjonctifs (« si Oswald n'avait pas tué Kennedy, alors quelqu'un d'autre l'aurait fait »). Elle montre la voie vers une analyse des conditionnels indicatifs à la manière de l'analyse proposée des contrefactuels. On n'a pas le temps de l'explorer dans ce travail.

contexte et que l'on définira dans §5.1.5, tiendra compte de cette structure. Un contexte est une version vague du cadre, qui respecte la structure dans la frontière imprécise entre prémisses et notion de conséquence. Il est vague, mais seulement selon une certaine dimension, pour ainsi dire.

Si l'on s'arrête sur l'imprécision de la frontière entre notion de conséquence et prémisses, et sur le vague qu'elle implique, c'est en raison de l'importance de ce caractère vague pour la pensée de la dynamique. Comme on l'a vu ailleurs (§2.3), le vague pourrait être un apport fondamental pour une pensée féconde de la dynamique, dans la mesure où elle permet la constitution d'une topologie et par conséquent une notion de continuité qui pourrait s'appliquer à la dynamique. Cette même stratégie sera employée dans le cas de contrefactuels, où la question est celle de la dynamique de l'apport pragmatique. Comme on discutera dans §5.1.5, la définition du contexte impliquera tout de suite une topologie, susceptible à représenter leur dynamique comme continue, et donc à étayer une théorie mûre de cette dynamique.

5.1.4 La frontière imprécise et arbitraire et le rapport entre cadres

Où l'on définit deux opérations entre des cadres qui correspondent au fait de transformer un aspect de la notion de conséquence en prémisses – la « distension » –, et inversement de transformer une prémisses en aspect de la notion de conséquence – la « restriction ». Le caractère arbitraire et imprécise de la frontière entre prémisses et notion de conséquence se conçoit en termes de ces opérations : il indique qu'il n'est pas toujours possible, pour un contrefactuel dans une situation donnée, de décider qu'un cadre est pertinent pour l'évaluation du contrefactuel, plutôt que certains de ses restrictions ou ses distensions.

Il est question ici du constat goodmanien (repris dans §4.2.2) d'une frontière « imprécise et arbitraire » entre la notion de conséquence et les prémisses. Une relation de conséquence qui joue un rôle dans une analyse pourrait aussi bien avoir servi de prémisses dans une autre analyse, et une prémisses, vue sous un autre angle, pourrait également avoir été une conséquence selon une (autre) notion de conséquence. D'où deux cas de « transferts » : pour une analyse particulière d'un contrefactuel (reposant sur une donnée d'une notion de conséquence et des prémisses particulières), on pourrait imaginer une autre analyse où ce qui était auparavant une prémisses a le statut d'une conséquence, ou autrement où ce

qui était naguère une conséquence prend le statut de prémisses. Par conséquent, l'imprécision dans la frontière entre prémisses et notion de conséquence indique que *l'on ne peut pas toujours trancher entre les analyses* (et les données de prémisses et de notions de conséquence sur lesquelles elles s'appuient) *qui sont reliées par de tels « transferts »*. Le but de cette section est de caractériser et de comprendre ces genres de « transferts ».

Dans la perspective présentée ci-dessus, ces « transferts » correspondent effectivement à deux opérations sur les cadres. La première est l'opération de *restriction* (Définition 5.1.10 à la fin du paragraphe) :

Soit (\mathbf{B}, E) un cadre, avec I l'ensemble d'énoncés atomiques qui engendrent l'algèbre d'interprétation de \mathbf{B} et soit A un énoncé de E . La restriction (par A) est le cadre (\mathbf{B}', E') , où

- *Si A appartient à I , alors \mathbf{B}' est l'algèbre interprétée obtenue à partir de \mathbf{B} en enlevant l'énoncé A . Si A n'appartient pas à I , alors \mathbf{B}' est l'algèbre interprétée qui a le même langage local que \mathbf{B} , mais où l'énoncé A est équivalent au faux.*
- *E' est l'ensemble obtenu à partir de E en enlevant A .*

La deuxième opération est l'opération inverse, que l'on appelle *distension* (Définition 5.1.11) :

Un cadre (\mathbf{B}, E) est une distension (par A) d'un cadre (\mathbf{B}', E') si et seulement si (\mathbf{B}', E') est une restriction (par A) de (\mathbf{B}, E) .

Dans l'opération de restriction, il s'agit de fixer un énoncé du langage local comme équivalent au vrai ou faux¹⁹, de telle sorte que la structure logique locale (c'est-à-dire l'algèbre interprétée) se trouve modifiée pour « incorporer » les nouveaux aspects structurels que cet énoncé exprime. Si l'on prend la restriction par $A \wedge \neg B$, alors B devient conséquence de A dans la structure logique obtenue par la restriction ; pour ainsi dire, l'aspect structurel exprimé par $A \wedge \neg B$ est *imposé* sur la structure logique locale (l'algèbre interprétée). La différence entre les deux cas de la définition est le sort de l'énoncé par lequel on fait la restriction : dans le premier cas, il est carrément enlevé de la structure logique et de son langage local, alors que dans le dernier cas, il reste dans la structure et dans le langage, quoique comme énoncé contradictoire (équivalent au faux) de l'algèbre. Dans le premier cas, le langage local est rétréci, dans le dernier cas, le langage est préservé. À l'inverse, dans l'opération de distension, on prend un énoncé qui est d'abord équivalent au vrai ou au faux, et on « défait » cette

¹⁹Si A est fixé comme équivalent au faux, $\neg A$ est fixé comme équivalent au vrai.

équivalence ; en conséquence, on ôte de la structure la relation ou l'aspect exprimé par cet énoncé. Si B est conséquence de A ($A \leq B$), alors $\neg A \vee B$ est équivalent au vrai dans la structure logique ; donc, si $\neg A \vee B$ n'est plus une équivalence au vrai après une distension, alors la relation $A \leq B$ ne vaut plus. Entre les deux cas de la définition, il y a le même genre de différence : dans le premier cas, on *ajoute* au langage local un énoncé qui en était auparavant absent²⁰, alors que dans le deuxième cas, puisque l'énoncé était toujours dans l'algèbre mais équivalent au vrai ou faux, on ne change pas le langage local. En bref, la restriction rend l'analyse du contrefactuel moins *raffinée* en rabattant une prémisse sur un aspect de la structure de l'algèbre, et donc sur un aspect de la notion de conséquence (§5.1.1) ; à l'inverse la distension *explicite* un aspect de la conséquence en l'ajoutant comme prémisse.

L'exemple du contrefactuel « si ce morceau de beurre était à 70°C, alors il fondrait » et d'une analyse qui prend comme seule prémisse « tout beurre fond à 70°C » servira à illustrer ces opérations. En écrivant A pour « ce morceau de beurre est à 70°C », B pour « il fond » et C pour « tout beurre fond à 70°C », on commence avec le cadre consistant en l'algèbre de Boole \mathbf{B} engendrée par quatre éléments, A , B , C , et R (reste), avec $B \equiv (A \wedge C) \vee R$ (le beurre fond si et seulement si tout beurre fond à 70°C et il est à 70°C, ou s'il y avait une autre raison qui la faire fondre), et l'ensemble co-indépendant $\{A, C, R\}$.

On obtient un autre cadre – et une autre analyse qui s'en sert – en enlevant la prémisse C (« tout beurre fond à 70°C ») et en le considérant comme aspect de la conséquence. Dans cette transition, on « convertit » une prémisse – un énoncé de l'algèbre – dans un aspect de la notion de conséquence – une propriété de la structure (§§4.2.1 et 5.1.1). Il s'agit de se mettre dans le cadre (\mathbf{B}', E') , où C n'est plus une prémisse, mais où son rôle est *englouti* dans la structure logique, en l'occurrence dans la structure de l'algèbre \mathbf{B}' . \mathbf{B}' est engendrée par trois éléments, A , B , et R , avec $B \equiv A \vee R$, E' devient $\{A, R\}$, en telle sorte que le contrefactuel est validé en raison au fait que B est conséquence de A dans \mathbf{B}' ($A \leq B$), c'est-à-dire grâce à la notion de conséquence seule²¹. La transition renvoyant (\mathbf{B}, E) à (\mathbf{B}', E') est un exemple de l'opération de restriction.

À l'inverse, dans la transition de (\mathbf{B}', E') à (\mathbf{B}, E) , le rapport structurel

²⁰L'énoncé est absent, mais s'il *était* présent, par la structure de l'algèbre, il *aurait été* équivalent au vrai ou au faux, de sorte que les considérations précédentes s'appliquent dans ce cas.

²¹Certains diraient peut-être qu'entre les deux algèbres les éléments A , B et surtout R prennent *intuitivement* des sens différents : si c'est le cas, c'est parce que, entre les algèbres, différents énoncés sont *présents*. Voir les remarques de §5.1.1 touchant à la localité du langage « interne » à l'algèbre interprétée.

entre A et B est transformée en un énoncé de l'algèbre non équivalent au vrai, à savoir C , lequel est pris comme prémisses. Entre (\mathbf{B}', E') et (\mathbf{B}, E) un aspect de la notion de conséquence est devenue prémisses : c'est l'explicitation qui a lieu lors d'une *distension*.²²

Alors que, comme l'a remarqué Goodman, la frontière entre les prémisses et la notion de conséquence est imprécise et arbitraire, chaque cadre représente un ensemble fixe de prémisses et une notion de conséquence fixe, c'est-à-dire une ligne de partage précise. Or, on aurait tendance à comprendre la frontière imprécise comme le fait que, pour une frontière préconisée, on peut absorber une prémisses dans la notion de conséquence et expliciter un aspect de la conséquence comme prémisses pour obtenir une frontière également tenable. Autrement dit, il y a imprécision parce que, pour un cadre proposé, certaines de ses restrictions et de ses distensions sont également « plausibles » comme « données pragmatiques » pour l'analyse. Ce qui suggérerait la stratégie qui consiste à rendre compte de la frontière imprécise en termes d'une collection de cadres qui sont des restrictions et des distensions les uns des autres. C'est la notion de contexte définie et développée dans le paragraphe suivant.

Pour définir rigoureusement l'opération de restriction, il faut définir préalablement l'opération de rétrécissement sur une algèbre interprétée.

Définition 5.1.9 (Le rétrécissement). Soit $\mathbf{B} = (B_I, B, q)$ une algèbre interprétée et A un élément de \mathbf{B} où $A \in I$ et $q(A) \in q(B_{I \setminus \{A\}})$. Le *rétrécissement* de \mathbf{B} par A est l'algèbre \mathbf{B}' tel que :

$$\mathbf{B}' \left\{ \begin{array}{ccc} B_{I \setminus \{A\}} & \xrightarrow{c} & B_I \\ \downarrow q|_{B_{I \setminus \{A\}}} & & \downarrow q \\ B & \xrightarrow{\simeq} & B \end{array} \right\} \mathbf{B}$$

Autrement dit, \mathbf{B}' a comme algèbre d'interprétation l'algèbre engendrée par $I \setminus \{A\}$, comme algèbre de base la même algèbre B avec la restriction de l'application de quotient de \mathbf{B} entre elles. C'est bien défini parce que l'image de $A \in B_I$ est dans $q(B_I)$.

La restriction peut maintenant se définir de la manière suivante.

²²Puisque l'énoncé C n'est pas équivalent à une combinaison logique d'autres énoncés de (\mathbf{B}, E) , il doit être traité comme un énoncé de I , comme on le fait dans le texte, et donc, selon la définition de restriction, enlevé du langage dans la transition à (\mathbf{B}', E') . Il s'agit donc du premier cas dans la définition de la restriction et de la distension. Pour un exemple du dernier cas, considère une algèbre (\mathbf{B}'', E'') , engendrée par A, B, P, Q , et R avec $B \equiv (A \wedge (P \vee Q)) \vee R$, avec ensemble co-indépendant $\{A, P \vee Q, R\}$ (par exemple, P est « le beurre a composition chimique γ » et Q est « le beurre a composition chimique δ », où il fondrait dans les deux cas). Si l'on restreint par $P \vee Q$, on ne peut pas l'enlever du langage local.

Définition 5.1.10 (La restriction). Soit (\mathbf{B}, E) un cadre et A un élément de E ($(\mathbf{B} = (B_I, B, q))$). Le *quotient* de \mathbf{B} par A est l'algèbre interprétée $\mathbf{B}/(\mathbf{A})$, avec la même algèbre d'interprétation que \mathbf{B} , mais, comme algèbre de base le quotient de B par $q(A)$ (Définition 5.1.4 dans §5.1.1). Le diagramme suivant commute :

$$\mathbf{B} \left\{ \begin{array}{ccc} B_I & \xleftarrow{=} & B_I \\ \downarrow q & & \downarrow q' \\ B & \xrightarrow{q''} & B/(q(A)) \end{array} \right\} \mathbf{B}/(\mathbf{A})$$

La *restriction* de (\mathbf{B}, E) par A est le cadre $(\mathbf{B}', E \setminus \{A\})$ où

Si $A \in I$: \mathbf{B}' est le rétrécissement de $\mathbf{B}/(\mathbf{A})$ par A ;

Si $A \notin I$: \mathbf{B}' est $\mathbf{B}/(\mathbf{A})$.

Le diagramme pour le second cas est celui ci-dessus ; le diagramme pour le premier cas est

$$\mathbf{B} \left\{ \begin{array}{ccccc} B_I & \xleftarrow{=} & B_I & \xleftarrow{\supset} & B_{I \setminus \{A\}} \\ \downarrow q & & \downarrow q' & & \downarrow q'|_{B_{I \setminus \{A\}}} \\ B & \xrightarrow{q''} & B/(q(A)) & \xleftarrow{\simeq} & B/(q(A)) \end{array} \right\} \mathbf{B}'$$

Définition 5.1.11 (La distension). (\mathbf{B}, E) est une *distension* de (\mathbf{B}', E') (par A) si et seulement si (\mathbf{B}', E') est une restriction de (\mathbf{B}, E) (par A).

5.1.5 Définition du contexte

Où une notion de contexte est définie comme une collection vague de cadres reliés par des restrictions et des distensions. Cette notion était une analyse plus raffinée des contrefactuels. De plus, dans la mesure où le sens du vague à l'œuvre ressemble à la notion d'ouvert provenant de la topologie, cette notion de contexte pourrait permettre de concevoir la dynamique comme continue.

Plusieurs analyses des contrefactuels se permettent, en plus de la donnée « pragmatique » requise par l'analyse (notion de conséquence et prémisses, relation de similitude, fonction probabiliste appropriée à un moment passé ou hypothétique ; §4.2.1), une *couche pragmatique* supplémentaire, telle par exemple un vague dans la relation de similitude (§4.2.3). En revanche, on s'efforce ici de prendre comme apport pragmatique une structure qui comporte en lui *l'entière* de la portée pragmatique pertinente à un instant donné. Pour le dire autrement, on prend comme apport pragmatique le *contexte en entier* : aussi bien la structure des énoncés en jeu (l'algèbre interprétée), les prémisses (l'ensemble

co-indépendant) que le *vague* dans cette structure et ces prémisses. Il s'agit, dans un premier temps, de définir une structure qui servira à modéliser un tel contexte vague, en l'occurrence en tant qu'un certain type de collection vague de cadres. En outre, force est de prendre conscience du rapport, fortement souligné dans la première partie de ce travail (notamment §2.3²³), entre le vague et la *dynamique*. Dans un deuxième temps, on verra que la conception proposée des contextes permet de les considérer comme des *ensembles ouverts* d'une topologie, ce qui est, comme déjà argué dans §2.3, le premier pas vers une compréhension de la dynamique, et notamment une compréhension de la dynamique comme *continue*.

La frontière entre prémisses et notion de conséquence est imprécise ; les « transferts » de « prémisses » en « aspect de la conséquence » et d'« aspect de la conséquence » en « prémisses » correspondent aux restrictions et aux distensions des cadres. Un contexte qui prend en charge cette imprécision ou ce vague dans la ligne – cette impossibilité de décider s'il s'agit d'un cadre de sa restriction ou de sa distension – semble se définir comme suit.

Un contexte est une collection ouverte²⁴ \mathcal{C} de cadres telle qu'entre toute paire de membres de la collection, il y a une chaîne de cadres de \mathcal{C} telle que chaque paire de cadres consécutifs de la chaîne est reliée par une opération de restriction ou de distension.

Si l'on dit qu'une collection \mathcal{C} de cadres est *connexe* s'il existe, pour toute paire de cadres de \mathcal{C} , une chaîne de cadres de \mathcal{C} qui les relie de la manière décrite, alors on peut re-exprimer cette définition de manière suivante : *un contexte est une collection ouverte et connexe de cadres* (Définition 5.1.13).

Dans la suite, la notation \mathcal{C} désignera aussi bien le contexte (au sens « ordinaire ») que la collection de cadres (qui le modélise). L'analyse suivante des contrefactuels, plus approfondie que la précédente, en ce sens qu'elle prend comme apports pragmatiques les contextes, s'ensuit immédiatement.

(ALA). *Un contrefactuel est évalué comme vrai dans un contexte \mathcal{C} s'il est évalué comme vrai dans tous les cadres de \mathcal{C} .*

La qualification d'une collection comme « ouverte » reprend effectivement l'approche du vague esquissée dans §§2.3.1 et 2.3.3 de la première partie, qui rapporte le vague à la notion d'ouverture topologique, plutôt qu'à une collection de toutes les « précifications » possibles de l'aspect, de la propriété, de

²³La discussion de §2.3 s'exprime pour l'essentiel dans des termes généraux, de sorte qu'elle s'applique directement ici.

²⁴Au sens de §2.3.1.

l'expression ou de la collection vague, à la façon d'un Fine²⁵. Afin de rappeler l'opposition entre ces deux approches, on considère l'exemple de la mesure de la longueur d'une table, mesure qui compte toujours un élément d'imprécision, voire de « vague ». Le philosophe analytique (à la Fine) la conçoit de manière suivante : il y a une longueur l dont la table est clairement plus longue, et une longueur l' dont elle est clairement plus courte ; la mesure vague est représentée par l'ensemble de nombres réels entre l et l' (écrit (l, l')) – les longueurs *précises possibles* de la table. Evidemment, la difficulté tient aux valeurs de l et de l' ²⁶. A l'opposé, le mathématicien (dont l'approche est empruntée ici) procède ainsi : il y a des ensembles (l_i, l'_i) tels que la longueur réelle de la table appartient à chacun d'entre eux (elle est entre l_i et l'_i) et tels que chacun de ces ensembles caractérise une mesure plus ou moins exacte selon l'approche du philosophe. Or, plutôt que de se poser la question de *quel ensemble* est le bon pour cette mesure particulière, le mathématicien se pose la question de la *structure qui subsiste sur la collection de ces ensembles*. Là où le philosophe pose une ligne pour mettre d'un côté le certain et le clair, voire le structuré, et de l'autre côté le vague comme abîme sans structure aucune, le mathématicien ne s'occupe pas d'une telle ligne mais plutôt de la structure du clair, du vague et, de plus important, du rapport entre eux. En un mot, il ouvre la voie à la pensée de *la structure du vague*. Car, à la différence du philosophe, il estime qu'il y a de telle structure : *le vague est structuré*.

Il s'avère que, dans le cas présent, les collections de cadres décrites sont bel et bien des *ensembles ouverts d'un espace topologique* et ont donc une structure « sur » eux. En outre, les rapports de restriction et de distension entre cadres ont un sens structurel relatif à cette topologie, dans la mesure où tous les ensembles ouverts de la topologie peuvent être formés par des unions des ensembles ouverts qui sont connexes au sens introduit ci-dessus, c'est-à-dire dont les éléments sont reliés par des chaînes de restrictions ou distensions. Alors, les ensembles ouverts et connexes ont un statut spécial dans cette topologie ; on appelle de tels ensembles des « ensembles ouverts de base »²⁷. Le contexte prend

²⁵Ce genre d'approche, dont Fine (1975) représente un exemple typique, est discuté dans §2.3.1. Quant à l'autre théorie discutée dans §2.3.1, à savoir celle de Williamson, des remarques semblables s'imposent relativement à elle, dans la mesure où les techniques logiques employées par Williamson pour comprendre l'ignorance humaine relative à un aspect dit vague, en l'occurrence les techniques offertes par la logique modale, ressemblent aux techniques impliquées dans la modélisation finienne du vague lui-même ; voir la note 41 de §2.3.1.

²⁶Les difficultés, les complexités et les conséquences à l'égard de la question du vague d'ordre supérieur qui s'accrochent à cette approche du vague sont discutées dans Fine (1975, §5) et Williamson (1994).

²⁷La conception d'ensemble ouvert « de base » (*basic open set*), dont tout autre ensemble

donc un sens précis : il est une collection *ouverte de base* (c'est-à-dire ouverte et connexe) au sens de cette topologie (Observation 5.1.1).

Il est difficile de surestimer l'importance du caractère ouvert ou vague du contexte comme une structure sur les cadres. Non seulement ceci balaie la « fausse question » de la ligne entre les aspects clairs et vagues du contexte : pour rendre compte de la frontière imprécise et arbitraire entre prémisse et notion de conséquence, par exemple, il faut comprendre les rapports entre les différentes collections des différents emplacements de cette frontière (différents cadres), non pas choisir une collection de préférence aux autres. Mais de plus, la notion d'ouverture (et d'ailleurs de connexité) est un apport nécessaire pour une théorie de la dynamique des contextes, et notamment pour une conception continue de la dynamique (§2.3.3) ; on aura l'occasion de s'attarder quelque peu sur cette question dans §5.2.3.

Avec la définition de la notion de contexte, et l'énonciation de l'analyse **(ALA)**, la présentation de la théorie de contrefactuels s'achève. Reste à l'évaluer et à considérer ces conséquences. Ce sont les tâches des deux prochaines sections respectivement.

Définition 5.1.12 (Connexité). Une collection \mathcal{C} de cadres est connexe, si, $\forall (\mathbf{B}, E), (\mathbf{B}', E') \in \mathcal{C}$,

$$\exists (\mathbf{B}, E) = (\mathbf{B}_0, E_0), (\mathbf{B}_1, E_1), \dots, (\mathbf{B}_{n-1}, E_{n-1}), (\mathbf{B}_n, E_n) = (\mathbf{B}', E')$$

tel que $\forall i, (\mathbf{B}_{n-1}, E_{n-1})$ est restriction ou distension de (\mathbf{B}_n, E_n) .

Définition 5.1.13 (Contexte). Un contexte \mathcal{C} est une collection connexe de cadres.

(ALA). Pour un contexte \mathcal{C} ,

$$\vdash_{\mathcal{C}} A \Rightarrow B \text{ ssi } \forall (\mathbf{B}, E) \in \mathcal{C}, \vdash_{(\mathbf{B}, E), v} A \Rightarrow B^{28}$$

La définition de contexte implique une topologie.

ouvert pourrait être formé par les unions, est bien connu en topologie (voir par exemple Sutherland (1975, §3.2)). Dans le cas présent, la topologie sur les cadres est la topologie discrète, mais la restriction d'intérêt aux ensembles ouverts « de base » – c'est-à-dire aux collections qui sont *connexes* – garantit que la mise en jeu de cette topologie n'est pas sans effet : dans les considérations ayant trait à la continuité par exemple, la contrainte de connexité n'est pas triviale (voir §5.2.3). Par ailleurs, la notion que l'on appelle ici « connexité » est plus forte que la notion de « connexité » dans la topologie (Sutherland, *op cit.*, Ch. 6).

²⁸On suppose une valuation « commune » pour tous les cadres du contexte (v). Autrement, on pourrait étendre facilement les opérations de restriction et de distension aux valuations, et remplacer « cadre » par « cadre avec valuation » partout dans la discussion. Voir également la note 16.

Observation 5.1.1 (Topologie). Soit \mathcal{T} la topologie sur la collection de cadres, dont les ensembles ouverts sont les O tel que

$$O = \bigcup_{O_i \in Conn} O_i$$

où $Conn$ est la collection de collections connexes de cadres.

\mathcal{T} est la topologie discrète. Par ailleurs, $Conn$ est une collection d'*ensembles ouverts de base* pour cette topologie, puisque $\forall O_1, O_2 \in Conn, O_1 \cap O_2 \in Conn$ ²⁹.

5.2 La théorie face aux défis

Une théorie des contrefactuels a été proposée ; il reste à l'évaluer. C'est l'objectif de cette section. Pour mener cette évaluation, il s'agit de s'appuyer sur la liste de *desiderata* auprès d'une théorie des contrefactuels établie dans le chapitre précédent (voir §4.3.4 pour le résumé). Grossièrement, cette liste comporte deux sortes de défis : d'une part, les défis « internes », référant aux propriétés des contrefactuels qui sont relatives à des situations individuelles, chacune mettant en jeu une seule valeur d'apport pragmatique (défis 1. à 4.) ; d'autre part, les défis « externes » ayant trait à plusieurs apports pragmatiques et aux rapports particuliers entre eux (défis 5. et 6.). En termes de la théorie proposée, les premiers défis mettent à l'épreuve la capacité de la notion de contexte de rendre compte des phénomènes qui concernent un contexte fixe, alors que les derniers défis interrogent la notion de contexte relativement à sa fécondité pour aborder les questions des rapports entre des contextes différents. Dans le premier paragraphe de cette section, on verra que du point de vue interne, la notion de contexte est largement suffisante : elle fournit une division entre les aspects « structurel » et « matériel » des contrefactuels de telle sorte que la frontière entre les deux est imprécise (défis 1. et 2.), elle respecte les phénomènes relatifs à l'indépendance, la pertinence (défi 3.), et elle rend compte du *hors sujet* (défi 4.). Les deux autres paragraphes de cette section seront consacrés aux défis externes, lesquels sont beaucoup plus épineux. Ils viseront à présenter ces problèmes sous le nouveau jour ouvert par la perspective proposée par l'analyse développée ci-dessus, à suggérer la fécondité des outils qu'elle a introduits, et à montrer jusqu'à quel point elle pourrait relever ces défis. Dans §5.2.2, on mettra en évidence quelques résultats techniques qui permettent de rendre compte du rapport entre les évaluations des contrefactuels apparentés dans des contextes différents (défi 5.) ; dans §5.2.3, on entame une considération du problème posé

²⁹Voir Sutherland (1975, §3.2).

par la dynamique des contextes – le *problème épineux des contrefactuels* –, en montrant que la dynamique des contextes est continue relativement à l'espace topologique défini dans §5.1.5.

Il sera question ci-dessous de la théorie pleine, exprimée en termes de contextes. Or la notion de contexte est complexe. Donc, pour des raisons tenant à la commodité, on travaillera à plusieurs moments avec l'analyse préliminaire (**ALP**) et avec des cadres « typiques » des contextes, plutôt que avec les contextes eux-mêmes.

5.2.1 Conséquences immédiates

Où l'on confirme que la théorie proposée satisfait aux quatre premiers défis de §4.3.4.

Sans même mettre en jeu le vague du contexte, l'analyse (**ALP**) et la notion de cadre satisfait d'ores et déjà à certains *desiderata* pour une théorie des contrefactuels.

- Elle réussit largement relatif au défi 1. de §4.3.4 : de rendre compte de la notion de conséquence et des prémisses requises pour une analyse des contrefactuels – ou autrement les aspects « matériels » et « structurels » – avec l'algèbre interprétée et l'ensemble co-indépendant respectivement. De plus, puisque, comme on l'a remarqué dans §4.2.1, les analyses des contrefactuels qui prennent cette forme évitent *a priori* les dits « sophismes contrefactuels », l'analyse proposée comporte certaines propriétés logiques de base que l'on exige souvent d'une théorie de contrefactuels.³⁰
- Elle rend compte de la co-indépendance des prémisses par rapport à l'antécédent, et d'ailleurs de la notion de co-indépendance compte tenu du conséquent (§4.2.3; défi 3. de §4.3.4), en termes de l'indépendance algébrique.

Par suite, elle évite le genre de problème identifié dans §4.2.3, touchant au fait que les analyses qui valident « si l'allumette avait été grattée, alors elle s'enflammerait » ($A \Rightarrow B$) risquent de valider également « si l'allumette avait été grattée, alors elle ne serait pas sèche » ($A \Rightarrow \neg G$; voir la discussion de l'exemple 1. de §5.1.3, à laquelle on emprunte la notation). S'il n'y avait pas d'exigence que les prémisses proviennent d'un ensemble co-indépendant, et que le conséquent soit équivalent à une combinaison logique des énoncés de cet ensemble co-indépendant, alors on

³⁰En outre, dans l'Annexe A.2, on montrera que, dans les limites imposées par la prise en compte du *hors sujet* et des phénomènes d'indépendance et de pertinence, cette analyse satisfait aux règles traditionnelles de la logique dite non-monotone.

pourrait faire l'analyse de ce contrefactuel avec l'algèbre \mathbf{B}_1 , engendrée par $\{A, B, G\}$ avec $A \wedge G \leq B$, algèbre dans laquelle $A \wedge \neg B \leq \neg G$ vaut également et donc où $A \Rightarrow \neg G$ est validé (prémisse : $\neg B$, c'est-à-dire, l'allumette ne s'enflamme pas). Or, l'exigence de la co-indépendance des prémisses et de l'antécédent compte tenu du conséquent, comprise dans le sens algébrique, oblige un recours à une algèbre comme \mathbf{B}_2 (voir l'exemple 1. de §5.1.3), comportant notamment un énoncé R , où il n'y a pas d'ensemble co-indépendant figurant A et B tel que G est équivalent à une combinaison logique de ses membres, et donc pas d'ensemble co-indépendant qui valide $A \Rightarrow \neg G$.

- On retrouve de manière tout à fait naturelle la distinction entre les semifactuels indépendants et semifactuels non indépendants (§4.2.3) : un semifactuel $A \Rightarrow P$ est indépendant s'il y a une combinaison logique des énoncés de l'ensemble co-indépendant qui est équivalent à P et qui ne figure pas A , il est non indépendant si toute combinaison logique équivalente à P figure A . La théorie est donc sensible à la pertinence de l'antécédent au conséquent (défi 3. de §4.3.4).

Par ailleurs, non seulement la théorie proposée *repère* les différents types de semifactuel, elle les traite de manières différentes. Alors que les semifactuels non indépendants sont évalués de la même manière que les contrefactuels (comme l'est d'ailleurs suggéré dans §4.2.3), l'analyse (**ALP**) n'attribue *aucune valeur de vérité aux semifactuels indépendants*. L'analyse est donc cohérente avec la complexité des semifactuels indépendants, tenant à l'hésitation sur leur valeur de vérité (c'est le problème de *verum ad quodlibet* de §4.2.3), ainsi qu'avec la position méthodologique adoptée dans ce travail, qui n'exige pas que l'on accorde une valeur de vérité à ces semifactuels, mais plutôt que l'on voit dans l'hésitation, dans *leur complexité même*, un phénomène à expliquer (§4.1.3). La théorie proposée affirme seulement *que* le conséquent est indépendant de l'antécédent, en laissant à celui qui l'emploie le soin de tirer quelle conclusion qu'il désire.

- La théorie proposée rend immédiatement compte du phénomène du *hors sujet* (§4.2.4 ; défi 4. de §4.3.4), grâce à la localité de l'algèbre interprétée : comme suffisamment discuté dans §5.1.1, une algèbre interprétée, et donc un cadre, ne comporte que certains énoncés du langage, si bien que les autres énoncés du langage y sont *hors sujet*.

Enfin, la mise en jeu du contexte dans l'analyse finale (**ALA**) permet de relever le dernier défi « interne ».

- L'analyse (**ALA**) rend compte du caractère imprécis et arbitraire de la

division entre notion de conséquence et prémisse (§4.2.2 ; défi 2. de §4.3.4), grâce à la définition de contexte (§5.1.5) de telle sorte que la distinction entre notion de conséquence et prémisse est vague.

Avant de se tourner vers les défis « externes », il convient de faire quelques remarques sur les contraintes strictes imposées par l'analyse de base **(ALP)**. Comme on l'a souligné dans §4.1.3, il s'agit ici de proposer une théorie qui vise à *comprendre*, ce qui n'implique pas qu'elle prend comme son but ultime d'accorder une valeur de vérité à tout contrefactuel. D'ailleurs, il y a apparemment des exemples, dont les cas de *verum ad quodlibet* et les semifactuels indépendants (§4.2.3), où il faut s'abstenir de forcer une valeur de vérité sur les contrefactuels, quitte à manquer d'autres phénomènes également importants. L'analyse **(ALP)** (et *a fortiori* **(ALA)**) s'inscrit foncièrement dans cette méthodologie. Elle accorde une valeur de vérité aux contrefactuels sous des conditions assez strictes (§5.1.3) ; autrement elle ne se prononce pas au sujet de leurs vérités, mais seulement fournit les autres propriétés qui empêchent une évaluation sûre (telle l'indépendance du conséquent par rapport à l'antécédent), laissant à l'utilisateur de l'analyse le choix de tirer quelle conclusion qu'il veuille. Elle est plutôt un *schème d'analyse* qu'une analyse exhaustive. Par conséquent, on peut employer le cadre général sans accepter les détails de **(ALP)** ; on peut étendre l'analyse à son gré en spécifiant des valeurs de vérités pour des contrefactuels qui ne satisfont pas à toutes les conditions qu'elle exige.

5.2.2 La comparaison de contextes

Où l'on comprend les rapports entre les évaluations des contrefactuels apparentés dans des contextes différents, en termes des relations de restriction et de distension entre les cadres de ces contextes. D'un côté, la communauté entre les évaluations qui accordent la même valeur de vérité au même contrefactuel, mais qui sont relatives à des contextes différents, est comprise en recourant à l'évaluation dans un cadre « minimal », dont tous les cadres des contextes respectifs sont des distensions. De l'autre côté, l'opposition entre les évaluations différentes d'un même contrefactuel se comprend en ramenant les cadres des différents contextes à des ensembles co-indépendants incompatibles dans une seule et unique algèbre interprétée « commune ».

Dans §4.3.2, on a mis en avant deux exemples de rapports entre des évaluations de contrefactuels apparentés dans des situations différentes qui sont

typiques du genre des rapports qu'une théorie de contrefactuel devra rendre compte. Il s'avère, dans les deux cas, qu'il y a des propriétés spécifiques des contextes et de leurs cadres qui permettent une compréhension de ces rapports. En les mettant en évidence, on illustre les possibilités offertes par la théorie proposée.

Le premier exemple a trait au contrefactuel « si ce morceau de beurre était à 70°C, alors il fondrait » : on s'attend à ce que ce contrefactuel, alors que vrai dans un grand nombre de situations, le soit pour à peu près la même raison dans toutes, à savoir qu'il y a une « loi » selon laquelle tout beurre fond à 70°C. En termes de cette analyse, il s'agit de comprendre le rapport entre les différents contextes (dans les différentes situations), c'est-à-dire les collections ouvertes et connexes de cadres. Techniquement, ce rapport semble être le suivant : il y a un « cadre minimal » dont tous les cadres dans les contextes appropriés sont des distensions (Définition 5.2.1). Ce cadre « minimal » a pour algèbre interprétée l'algèbre engendrée par $A =$ « le beurre est à 70°C », $B =$ « il fond », et R (le « reste ») telle que $B \equiv A \vee R$, et pour ensemble co-indépendant $\{A, R\}$. La dite « loi » selon laquelle tout beurre fond à 70°C est imbriquée dans la structure logique (de l'algèbre interprétée) en tant que relation de conséquence entre A et B ($A \leq B$). Or, d'un côté, toute distension de ce cadre comportera quelque chose qui ressemble à cette « loi », qu'il soit plus ou moins précise, explicite et facilement repérable. À l'inverse, il semble que tout cadre qui valide le contrefactuel en « raison » de quelque chose de l'ordre d'une « loi » selon laquelle B est conséquence par A (éventuellement en conjonction avec d'autres énoncés), serait une distension de ce cadre « minimal »³¹. Ainsi, la communauté entre les contextes qui valident le contrefactuel (« en raison » de cette « loi ») correspond au fait que leurs cadres sont des distensions de ce cadre « minimal ». Dans un certain sens, ce cadre « résumé » ce qui est commun d'entre eux³².

Le second exemple relevé dans §4.3.2 concerne le rapport entre les deux contrefactuels suivants.

- (E1) « Si la ville de New York était dans l'état de Georgie, alors la ville de New York serait dans le Sud »
- (E2) « Si la Georgie incluait la ville de New York, alors Georgie ne serait pas complètement dans le Sud »

³¹Certainement, en raison des contraintes de l'analyse (ALP) tout tel cadre comportera un « reste » qui serait l'image de R dans la distension.

³²Cela n'implique nullement que ce cadre *appartient* à tous les contextes, car la définition du contexte (§5.1.5) exige que seulement la collection de cadres soit connexe, non pas que la collection de cadres soit close sous la connexité.

Puisque ces deux contrefactuels sont contradictoires, il n'y a aucune possibilité de trouver un cadre « canonique » qui valide les deux et qui est « commun » aux différents contextes. Néanmoins, techniquement, le rapport entre les contrefactuels se comprend de manière suivante : il y a une algèbre interprétée commune, et deux ensembles co-indépendants différents de cette algèbre telle que le cadre consistant en l'algèbre et l'un des ensembles co-indépendants valide l'un des contrefactuels et le cadre consistant en l'algèbre et l'autre ensemble co-indépendant valide l'autre. Les rapports entre les contextes où les différents contrefactuels sont validés se comprennent en termes des rapports entre ces ensembles co-indépendants dans cette seule algèbre. L'existence d'une telle algèbre munie de tels ensembles co-indépendants est assurée par le résultat suivant (Corollaire 5.2.1 ci-dessous).

Pour deux cadres (\mathbf{B}_1, E_1) et (\mathbf{B}_2, E_2) , il y a une algèbre interprétée \mathbf{B}' et deux ensembles co-indépendants E'_1 et E'_2 de \mathbf{B}' telle que (\mathbf{B}', E'_1) est une distension de (\mathbf{B}_1, E_1) et (\mathbf{B}', E'_2) est une distension de (\mathbf{B}_2, E_2) .³³

Pour les deux contrefactuels de l'exemple (où l'on emploie, comme dans §4.2.3, A pour l'antécédent commun, B pour le conséquent de $(\mathbf{E1})$ et G pour la négation du conséquent de $(\mathbf{E2})$), $(\mathbf{E1})$ est évalué comme vrai dans le cadre (\mathbf{B}_1, E_1) où \mathbf{B}_1 engendrée par A, B, G et R avec $B \equiv (A \wedge G) \vee R$, et E_1 est $\{A, G, R\}$, et $(\mathbf{E2})$ est évalué comme vrai dans le cadre (\mathbf{B}_2, E_2) où \mathbf{B}_2 engendrée par A, B, G et R' avec $\neg G \equiv (A \wedge \neg B) \vee R'$, et E_2 est $\{A, B, R'\}$. Pour comprendre le rapport précis entre ces deux évaluations, relatives à ces deux cadres, on emploie l'algèbre \mathbf{B} engendrée par $\{A, B, G, R, R'\}$ avec $(A \wedge \neg G) \vee R \equiv B$ et $G \equiv (A \wedge \neg B) \vee R'$. Pour les deux ensembles co-indépendants $E'_1 = \{A, G, R\}$ et $E'_2 = \{A, B, R'\}$ de \mathbf{B} , (\mathbf{B}, E'_1) et (\mathbf{B}, E'_2) sont des distensions de (\mathbf{B}_1, E_1) et (\mathbf{B}_2, E_2) respectivement ; en ce sens \mathbf{B} est une algèbre « commune » apte à comparer les deux cadres et leurs évaluations. Le rapport entre les deux évaluations se comprend immédiatement en termes du rapport entre les deux ensembles co-indépendants E'_1 et E'_2 de \mathbf{B} : ils sont non seulement différents mais *incompatibles*, en ce sens que leur union n'est pas un ensemble co-indépendant³⁴. Voilà l'expression rigoureuse de la différence entre les évaluations : il s'agit de deux choix *incompatibles* de prémisses (E'_1 et E'_2) pour une même notion de conséquence (la structure de \mathbf{B}).

³³Dans cette formulation, on fait abstraction des ajouts de nouveaux énoncés à l'algèbre interprétée qui ne changent pas la structure de l'algèbre de base, c'est-à-dire les « rétrécissements » de §5.1.4 (Définition 5.1.9). Pour les détails précis, voir les résultats à la fin du paragraphe.

³⁴Voir §5.1.2 sur l'incompatibilité des ensembles co-indépendants.

Cette considération de deux cadres qui évaluent les contrefactuels de manières différentes s'étend sans problème à la comparaison de deux contextes. De fait, les cadres considérés (\mathbf{B}_1, E_1) et (\mathbf{B}_2, E_2) sont les cadres « minimaux » relatifs aux deux contrefactuels, au sens précisé ci-dessus, de telle sorte que l'on peut faire le rapport entre n'importe quelle paire de contextes dont l'un valide **(E1)** et l'autre valide **(E2)** en passant par ces cadres « minimaux ». Les contextes sont « inverses » dans la mesure où les distensions de ces cadres « minimaux » ayant pour algèbre \mathbf{B} comportent des ensembles co-indépendants « inverses ».³⁵

En somme, le cadre théorique proposé possède les moyens pour rendre compte des rapports entre les évaluations des contrefactuels apparentés dans des situations différentes. Il s'agit de la possibilité de trouver des structures logiques locales (algèbres interprétées) « communes » ou « minimales », qui entretiennent des cadres auxquels les cadres des contextes en question sont reliés par des distensions ou des restrictions, et de comparer les ensembles co-indépendants appropriés dans ces algèbres. Cette discussion sert à indiquer la fécondité de la conception développée ci-dessus pour penser les contrefactuels et leurs subtilités.

Définition 5.2.1 (Cadre minimal). Un *cadre minimal* d'une collection de contextes $(\mathcal{C}_i)_{i \in I}$ est un cadre (\mathbf{B}, E) tel que

$$\forall i \in I, \forall (\mathbf{B}_j^i, E_j^i) \in \mathcal{C}_i, (\mathbf{B}_j^i, E_j^i) \text{ est une distension de } (\mathbf{B}, E)$$

Le résultat employé dans la considération des contrefactuels **(E1)** et **(E2)** s'appuie sur la définition suivante de produit, comme une sorte de « fusion » de deux algèbres interprétées.

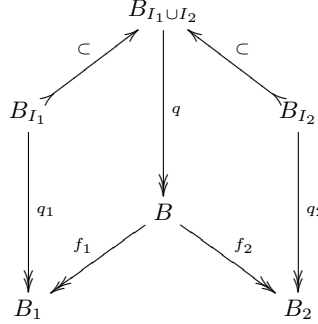
Définition 5.2.2 (Produit). Le *produit* de deux algèbres interprétées $\mathbf{B}_1 = (B_{I_1}, B_1, q_1)$ $\mathbf{B}_2 = (B_{I_2}, B_2, q_2)$ est l'algèbre interprétée $\mathbf{B}_1 \times \mathbf{B}_2 = (B_{I_1 \cup I_2}, B, q)$, où $B = B_{I_1 \cup I_2} / (\mathcal{F}_1 \cup \mathcal{F}_2)$ et q est l'application de quotient³⁶.

³⁵Par ailleurs, on peut aussi rendre compte d'une troisième intuition relative à cette paire de contrefactuels, selon laquelle ni l'un ni l'autre n'est vrai. (Comme remarqué dans la note 35, le sens de cette intuition est sujet à un débat entre Lewis (1973b, §3.4), qui l'identifie comme un aspect « logique » de contrefactuels, et Stalnaker (1980) qui y voit un aspect relevant de la « pragmatique » ; il n'est besoin d'entrer dans ce débat ici.) Il s'agit d'employer l'algèbre \mathbf{B} avec l'ensemble co-indépendant $\{A, R, R'\}$, lequel n'évalue aucun des contrefactuels (selon l'analyse **(ALP)**). Puisque $\{A, B, R, R'\}$ et $\{A, G, R, R'\}$ sont co-indépendants, cela est fidèle à l'idée que l'on peut décider en faveur de l'un ou de l'autre contrefactuel en « rendant plus précis » les prémisses (en l'occurrence l'ensemble co-indépendant $\{A, R, R'\}$).

³⁶Voir la remarque suivant la Définition 5.1.3 de §5.1.1 pour la définition des \mathcal{F} et la conception de l'algèbre interprétée comme un quotient.

Cette opération de produit a la propriété importante suivante³⁷.

Proposition 5.2.1. *Pour une paire quelconque d'algèbres interprétées \mathbf{B}_1 et \mathbf{B}_2 , chaque \mathbf{B}_i est un rétrécissement d'un quotient de l'algèbre $\mathbf{B}_1 \times \mathbf{B}_2$ (cf. Définition 5.1.9). Autrement dit, il y a des applications de quotient $f_i : B \rightarrow B_i$ telles que le diagramme suivant commute.*



D'après les Définitions 5.1.10 et 5.1.11, on a le Corollaire suivant.

Corollaire 5.2.1. *Soit (\mathbf{B}_1, E) un cadre, et \mathbf{B}_2 une algèbre interprétée. Alors $(\mathbf{B}_1 \times \mathbf{B}_2, E')$ est le rétrécissement d'une distension de (\mathbf{B}_1, E) (où $E' \subset \mathbf{B}_1 \times \mathbf{B}_2$ est l'image de $E \subset \mathbf{B}_1$ dans la distension).*

5.2.3 La dynamique de contextes

Où l'on discute un exemple où l'évaluation d'un contrefactuel est intuitivement conséquence du discours précédent ; autrement dit, l'évaluation du contrefactuel se comprend en termes de la dynamique du contexte. On montre que, dans cet exemple, la notion de contexte définie ci-dessus permet de représenter la dynamique du discours comme continue. Alors que ceci ne constitue pas une démonstration infaillible, c'est néanmoins une indication que la conception de contrefactuels proposée ici a de larges chances de relever le premier défi posé par la dynamique : de la penser comme continue.

Le défi le plus difficile pour une théorie des contrefactuels consiste à comprendre la dynamique de leur apport pragmatique (§4.3.3, défi 6. de §4.3.4). Chaque théorie des contrefactuels adopte ou propose une conception d'apport pragmatique susceptible à supporter une évaluation de contrefactuels (relation de similitude, notion de conséquence et prémisses, fonction probabiliste, dans

³⁷On ne reproduit pas la démonstration ici. Elle consiste essentiellement à prendre pour f_1 l'application de quotient pour le quotient par $F = \{x|_1 \triangle x : x = q(A), A \in I_2 \setminus I_1\}$, où, pour $x \in B$, $x|_1$ est l'élément le plus grand qui est dans l'image de B_{I_1} sous q et qui est au dessous de $(\leq) x$.

le cas de la théorie proposée ici, contexte au sens de §5.1.5) ; il s'agit de savoir dans quelle mesure cette conception offre une compréhension de la dynamique de l'apport pragmatique. Il ne suffit pas de présenter les aspects que l'on observe intuitivement auprès des transitions dynamiques dans des termes plus rigoureux : il faut de plus que cette formulation soit révélateur, qu'elle contribue à la compréhension de ces transitions dynamiques. La première étape qui doit être franchie par toute représentation de l'état instantané (en l'occurrence, de l'apport pragmatique) qui prétend permettre une pensée de la dynamique a été souligné à plusieurs reprises antérieurement (§§2.3 et 5.1.5) : il doit représenter la dynamique comme *continue*. Une compréhension d'une transition dynamique doit au moins rendre compte du genre de « chevauchement » ou « proximité » entre les états antérieur et ultérieur.

On a su proposer une notion d'apport pragmatique – le contexte – qui induit naturellement une topologie et donc une notion de continuité (§5.1.5). Il n'est pas question de démontrer rigoureusement que la dynamique *effective* de contextes est *toujours* continue selon cette notion de continuité ; cela semble d'ailleurs une question empirique plutôt qu'une question qui se décide *a priori*. On peut cependant arguer informellement que la dynamique est continue. C'est la tâche de ce paragraphe : montrer la continuité de la dynamique par la considération d'un exemple. Au passage, par le fait même de mettre en œuvre les outils techniques développés auparavant, on aura l'occasion d'illustrer le champ, l'importance et la fécondité de leurs applications possibles.

Il s'agit en l'occurrence du texte suivant.

La veille de la guerre civile, quelques états confédérés ont songé à l'annexion des territoires de l'Union, et au problème de comment répartir ces territoires parmi eux. Une chose était certaine : Georgie voulait la ville de New York. Que c'est drôle à penser, les possibilités qu'on trouve dans l'histoire ! Georgie, un symbole du Sud ; mais si elle incluait la ville de New York, alors elle ne serait pas complètement dans le sud.

Conformément à l'image générale du contexte comme un champ où figurent *tous* les énoncés et les aspects plus ou moins *vaguement* en jeu avec les rapports structurels appropriés entre eux (§5.1), laquelle sous-tend la notion de contexte proposée dans §5.1.5, cette notion s'applique non seulement aux moments où il s'agit de contrefactuels, mais tout au long de la lecture ou la considération de ce texte³⁸. La question est de savoir le rapport entre les contextes à des instants différents ; comme remarqué dans §4.3.3, c'est en répondant à cette question que

³⁸Dans ce sens, elle rejoint à la notion de contexte débattue dans la littérature ; voir par exemple Stalnaker (1999, Partie 1) ou Lewis (1979c).

l'on arrive à expliquer pourquoi (ou même : comment) on a tendance à évaluer le contrefactuel à la fin comme vrai.

Il n'est pas difficile de décrire les rapports entre les contextes consécutifs à l'œuvre dans la lecture de ce texte : entre le contexte approprié au premier énoncé, qui figure des énoncés au sujet des guerres, de l'annexion de territoires, de la guerre civile américaine, et le contexte approprié au second énoncé, figurant les détails plus précis relatifs à New York, à la Géorgie, de leurs côtés dans la guerre, certains énoncés sont ajoutés (la Géorgie est un état confédéré, par exemple), certains sont gardés (les territoires conquises par les états confédérés seraient partagées parmi ces états), et certains disparaissent du contexte (les états confédérés ont considérés la partition des territoires conquises, peut-être). De même, entre le contexte approprié au second énoncé et celui approprié au dernier énoncé³⁹, certains énoncés au sujet de la guerre et de l'annexion sont peut-être sortis, et d'autres énoncés, ayant trait notamment à la géographie de la Géorgie et de New York, sont entrés. *Intuitivement*, on comprend l'évaluation positive du contrefactuel dans le dernier énoncé en comprenant la démarche dynamique qui a mené à ce contexte. La difficulté consiste à concevoir la démarche d'une manière susceptible à apporter un brin de précision et de compréhension à ces intuitions floues mais justes. Pour que la notion de contexte proposée ici soit à la hauteur de cette tâche, pour qu'elle soit plausible et utile comme manière de représenter la discussion, et de considérer le contrefactuel à la fin, il faut qu'elle représente cette démarche comme continue.

Elle serait continue si, entre deux contextes proches dans le temps, il y avait une suite de contextes telle que chaque paire de contextes adjacents dans la suite se chevauchent. D'après la définition de contextes comme des collections ouvertes et connexes de cadres (§5.1.5), cela est équivalent à dire que, entre deux contextes (collections ouvertes et connexes de cadres) proches dans le temps, il y a une suite de collections ouvertes et connexes de cadres telle que chaque paire de collections adjacentes se chevauchent, c'est-à-dire, telle que chaque paire de collections adjacentes ont des cadres en commun⁴⁰. Puisque la connexité de ces collections implique que tous les cadres sont reliés par des chaînes de restrictions et distensions, il suffit de montrer que, pour un cadre d'un contexte

³⁹ On fait abstraction du troisième énoncé, dont l'effet est largement rhétorique.

⁴⁰ Comme remarqué dans la note 27, puisque la topologie est effectivement discrète (Observation 5.1.1) – c'est-à-dire toute collection est ouverte –, la conception de continuité comme l'exigence d'une suite de collections ouvertes (non nécessairement connexes) est trop faible pour être intéressante. En revanche, il est moins trivial de savoir s'il est possible de penser la dynamique en termes des collections qui sont également connexes, dans la mesure où il faudrait dans ce cas des relations appropriées de restriction and de distension entre les cadres.

et un cadre de l'autre, il existe une chaîne de cadres reliés par des restrictions ou des distensions.

La transition entre un contexte approprié au second énoncé du texte ci-dessus et un contexte approprié au dernier énoncé sert d'exemple. Un cadre « typique » du contexte relatif au second énoncé est (\mathbf{B}_1, E_1) , avec \mathbf{B}_1 engendrée par les énoncés « la Géorgie veut New York » (P_1), « la Géorgie peut réclamer des territoires conquis » (P_2), « New York fait partie des territoires annexés » (P_3), et « la Géorgie annexe New York » (D), avec $D \equiv P_1 \wedge P_2 \wedge P_3$, et $E_1 = \{P_1, P_2, P_3\}$. Un cadre « typique » du contexte approprié au dernier énoncé est (\mathbf{B}_2, E_2) , avec \mathbf{B}_2 engendrée par les énoncés « la Géorgie est dans le Sud » ($\neg G$), « New York n'est pas dans le sud » ($\neg B$), « l'ajout de territoire modifie la géographie d'un état » (P_4), « la Géorgie annexe New York » (D) et « La Géorgie inclut New York » (A), avec $\neg G \equiv D \wedge P_4 \wedge \neg B$ et $A \equiv D$, et $E_2 \equiv \{A, P_4, B\}$ ⁴¹. Par le résultat mobilisé dans le paragraphe précédent (Corollaire 5.2.1), il y a une algèbre \mathbf{B} figurant tous les énoncés dans \mathbf{B}_1 et \mathbf{B}_2 ⁴² et deux ensembles co-indépendants de \mathbf{B} , $E'_1 = \{A, P_4, B\}$ and $E'_2 = \{P_1, P_2, P_3\}$ telle que (\mathbf{B}, E'_1) (respectivement (\mathbf{B}, E'_2)) est une distension de (\mathbf{B}_1, E_1) (respectivement (\mathbf{B}_2, E_2)). L'union de ces ensembles co-indépendants n'est pas co-indépendant; pour reprendre le terme de §5.1.2, ils sont incompatibles. Or, cette incompatibilité est conséquence uniquement du fait que A est une combinaison logique des éléments de E'_2 ⁴³, ou, pour le dire dans des termes introduits dans §4.2.3, il est « dépendant » de ou « déterminé » par E'_2 . Il s'ensuit que tout énoncé qui est équivalent à une combinaison logique des énoncés de E'_1 et E'_2 est également équivalent à une combinaison logique qui ne figure pas A ; A n'ajoute rien aux énoncés qui peuvent être formés comme combinaisons logiques à partir de E'_1 et E'_2 . De plus, l'ensemble obtenu en enlevant A de l'union de E'_1 et E'_2 , $\{P_1, P_2, P_3, P_4, B\}$, est co-indépendant. Ce fait témoigne d'une certaine « compatibilité » entre les cadres.

De fait, le cadre $(\mathbf{B}, \{P_1, P_2, P_3, P_4, B\})$ établit le lien entre (\mathbf{B}_1, E_1) et (\mathbf{B}_2, E_2) . La restriction par P_4 et B ⁴⁴ fournit le cadre (\mathbf{B}_1, E_1) ; la restriction par P_2 et P_3 donne le cadre (\mathbf{B}_2, E_2) ⁴⁵. Alors, il y a une chaîne, consistant

⁴¹ Alors que ces cadres peuvent être moins exact comme modèles des contextes, dans la mesure où l'on pourrait préférer des cadres avec des restes (§5.1.3), ils sont employés pour simplifier la discussion. Les conclusions de cette discussion s'appliquent également aux cadres avec restes. Voir la note 17.

⁴² Précisément : \mathbf{B} est le produit des deux algèbres (Définition 5.2.2), c'est-à-dire l'algèbre comportant tous les énoncés de \mathbf{B}_1 et \mathbf{B}_2 , avec les « équivalences appropriées » entre eux.

⁴³ Puisque $A = D = P_1 \wedge P_2 \wedge P_3$.

⁴⁴ Combiné avec des rétrécissements par A et G ; voir la Définition 5.1.9 de §5.1.4.

⁴⁵ Modulo le rétrécissement par P_1 .

en une distension et une restriction et passant par $(\mathbf{B}, \{P_1, P_2, P_3, P_4, B\})$, entre (\mathbf{B}_1, E_1) et (\mathbf{B}_2, E_2) . En outre, ce cadre intermédiaire rend manifeste la « raison » pour que le contrefactuel **(E2)** soit évalué comme vrai : puisque $\neg B$ (et non $\neg G$) est indépendant des données de l'histoire P_1 à P_4 , au sens où l'ensemble de ces énoncés est co-indépendant, et puisque A est impliqué par ces données, il pourrait être pris comme prémisses « fixe », indépendant de l'antécédent du contrefactuel (et de son conséquent). Ainsi, le contrefactuel est vrai, selon l'analyse **(ALP)**.

En résumé, des contextes ont été identifiés à des « collections ouvertes de base » (c'est-à-dire ouvertes et connexes) d'une topologie (§5.1.5) ; la continuité de la dynamique de contextes, définie en termes de cette topologie, consiste en le fait pour chaque paire de contextes proches dans le temps d'être reliés par une chaîne de contextes (« collections ouvertes de base ») dont chaque paire adjacente se chevauche. On a pu montrer, sur un exemple où l'évaluation d'un contrefactuel s'explique en termes de la dynamique antérieure de l'apport pragmatique, que la dynamique des contextes impliquée est effectivement continue au sens défini. Bien entendu, ceci ne permet pas de déduire que la représentation de la dynamique fournie par la notion de contexte introduite ici sera *toujours*, *hors la moindre doute* continue, si une telle démonstration est même possible. Pourtant, elle sert à indiquer que cette notion de contexte marche de manière générale, qu'elle est utile et féconde comme manière de concevoir la dynamique de l'apport pragmatique, et par conséquent que la théorie des contrefactuels proposée, reposant sur cette notion de contexte et l'analyse **(ALA)** qui s'exprime en termes d'elle, a fait d'importants progrès sur la question épineuse de la dynamique.

5.2.4 Résumé : Bilan de la théorie proposée

Relativement aux défis établis dans le Chapitre 4 (résumés et numérotés dans §4.3.4), la théorie proposée dans §5.1 compte les réussites suivantes.

1. La part non formelle requise pour l'analyse d'un contrefactuel est un contexte, une collection ouverte et connexe de cadres, où un cadre consiste en une algèbre interprétée et un ensemble co-indépendant. L'analyse du contrefactuel dans un cadre (et *a fortiori* dans un contexte) est de la forme générale **(AGC)** : un contrefactuel est évalué comme vrai si la conjonction de l'antécédent avec les membres de l'ensemble co-indépendant (les « prémisses ») a pour conséquence le conséquent, selon les rapports de conséquence instaurés par la structure de l'algèbre (la « notion de

conséquence »). Les propriétés générales de la forme d'analyse (**AGC**), dont notamment le fait d'éviter les dits « sophismes contrefactuels », sont héritées par l'analyse proposée ici.

2. Un contexte est une collection ouverte et connexe de cadres : entre chaque paire de cadres d'un contexte, il y a une chaîne de distensions et de restrictions. Puisqu'une transformation d'une prémisse en un aspect de la notion de conséquence est un exemple d'une restriction (et la transformation inverse est une distension), cette définition du contexte rend compte du caractère flou de la distinction entre les prémisses et la notion de conséquence impliquées dans la considération d'un contrefactuel.
3. L'indépendance algébrique qui vient naturellement avec l'algèbre caractérise l'indépendance constatée des prémisses par rapport à l'antécédent et par rapport à elles-mêmes ; l'exigence que le conséquent soit une combinaison logique des énoncés de l'ensemble co-indépendant assure l'indépendance « compte tenu » du conséquent. L'indépendance algébrique permet également d'identifier et de comprendre les phénomènes ayant rapport à la pertinence de l'antécédent par rapport au conséquent, dont les cas de *verum ad quodlibet*.
4. Un cadre ne comporte pas généralement tous les énoncés du langage ; de même pour un contexte. Les énoncés qui n'appartiennent pas au contexte sont *hors sujet*.
5. Les rapports entre des contextes différents sont compris par recours à des cadres qui sont reliés aux cadres des contextes par des distensions et des restrictions. Différentes propriétés de tels cadres et de ses ensembles co-indépendants correspondent aux différents rapports entre les contextes eux-mêmes.
6. Non seulement la notion de contexte qui constitue la pierre angulaire de la théorie proposée étaye une topologie et donc une notion de continuité, mais de plus, il y a de bonnes raisons pour penser que la dynamique des contextes est continue (selon cette notion de continuité). Ceci est un résultat important, dans la mesure où une conception continue de la dynamique est la première étape envers une compréhension approfondie de la dynamique.

Plus une problématique est difficile, plus une théorie qui l'accepte est ambitieuse, moins on en attend une résolution complète et finie. Cette théorie ne résout pas totalement les questions épineuses des contrefactuels, et notamment celles ayant trait à leur dynamique. Néanmoins, elle se montre capable de faire

des progrès importants, en introduisant un ensemble de concepts d'apparence utiles et féconds. Si elle n'arrive pas au bout du chemin, du moins elle fait les premiers pas. Elle montre une voie.

5.3 Conséquences de la théorie : l'aller et retour entre sémantique et pragmatique

La problématique des contrefactuels à laquelle a été consacré ce travail diverge nettement des problématiques traditionnelles. Plus notamment, on a rejeté, dans sa formulation, un appui excessif sur une distinction entre la « sémantique », ou la « logique », et la « pragmatique » (§4.3.1). En conséquence, à la différence de la question, souvent traitée dans la littérature, des propriétés « logiques » ou « sémantiques » des contrefactuels, il s'agit ici de la question générale des contrefactuels, relative aussi bien à leur côté pragmatique que sémantique : il s'agit du « bloc sémantique-pragmatique », pour ainsi dire.

La théorie proposée dans §5.1 porte la marque du questionnement auquel elle répond. En particulier, le côté « sémantique » – la structure des contextes et l'évaluation des contrefactuels en termes d'eux – et le côté « pragmatique » – la structure des rapports entre les différents contextes – s'imbriquent. Pour accoutumer un lecteur peu familier avec ce genre de flexibilité dans le rapport entre pragmatique et sémantique, et pour lui persuader que cette flexibilité n'a rien d'anormal, voire même qu'elle devrait être considérée comme un atout, il convient de montrer comment certains phénomènes ayant trait aux contrefactuels peuvent être compris de manière également suffisante en termes de la sémantique qu'en termes de la pragmatique : comme si, pour ces phénomènes, la question de savoir s'ils ressortent à la pragmatique ou à la sémantique n'a pas de sens. Dans le premier paragraphe de cette section, on se tâche d'illustrer ces possibilités sur un exemple concret qui occupe par ailleurs une place importante dans la littérature sur les contrefactuels.

Une conséquence supplémentaire de la minoration de la distinction entre sémantique et pragmatique, qui a été également remarquée dans §4.3.1, consiste en la perte d'importance des règles logiques. Ceux qui demeurent attachés à l'idée, et à la problématique, des règles logiques, peuvent être déçus par leur abandon. Ils n'ont pourtant rien à craindre : le fait de délaisser les règles logiques tient *uniquement* à la conviction que, en ce qui concerne la vraie question des contrefactuels, elles n'offrent pas de réponse suffisante. Il *n'implique aucune* quelconque impossibilité de rendre les prétendues règles logiques dans le cadre conceptuel proposé. Au contraire, les règles logiques peuvent correspondre

à trois aspects selon cette conception — des propriétés de la structure (interne) d'un contexte — des propriétés des transitions entre contextes — les propriétés d'une structure « globale » qui « résume », en quelque sorte, tous les contextes individuels. La conception des contrefactuels impliquée par la théorie s'avère donc être quelque chose de plus générale : elle n'est pas moins qu'*une conception de ce que c'est la logique*. Ce style de penser des questions souvent dites logiques, que l'on appelle la *logique locale*, s'applique à un vaste champ de questionnements : non seulement il convient aux problématiques qui se chargent d'aspects pragmatiques, comme celle dont s'occupe la présente partie de la thèse, mais il est également approprié aux traitements des systèmes et des questionnements logiques traditionnels. Du coup, il permet une compréhension approfondie des rapports entre ces problématiques différentes, et entre les théories différentes qui y répondent. Il s'agit, dans le dernier paragraphe de cette section, d'esquisser et de motiver brièvement quelques aspects de la logique locale.

5.3.1 Flexibilité de la partition entre sémantique et pragmatique

Comme l'est entraînée par la problématique posée à l'égard des contrefactuels, la théorie développée comporte une frontière flexible entre sa part « sémantique » (la structure des contextes) et sa part « pragmatique » (les rapports entre contextes). En particulier, divers phénomènes relatifs aux contrefactuels peuvent être analysés aussi bien comme « sémantiques » que comme « pragmatiques ».

Aucune distinction stricte entre un côté « sémantique » et un côté « pragmatique » des contrefactuels n'est supposée ici. Si l'on accorde à appeler « sémantiques » les aspects ayant égard à l'évaluation de contrefactuels dans un contexte fixe (les aspects appelés « internes » dans §5.2) et « pragmatiques » les aspects relatifs aux rapports entre des contextes différents (les aspects appelés « externes »), la théorie proposée admet des rapports complexes entre les côtés sémantique et pragmatique. De plus important, il y a des phénomènes relatifs aux contrefactuels qui sont susceptibles d'analyses « sémantiques » – dans un seul contexte – aussi bien que « pragmatiques » – en termes des transitions entre contextes –, de telle sorte que l'on ne saurait pas où les classer.

Une illustration frappante est l'exemple, dont Lewis s'est servi pour motiver sa propre analyse, des suites de contrefactuels, avec des antécédents de plus en plus précisés, de forme générale :

$$A_1 \Rightarrow B ; A_1 \& A_2 \Rightarrow \text{non } B ; A_1 \& A_2 \& A_3 \Rightarrow B ; \dots$$

Par exemple : « Si Otto était venu, la soirée serait animée; or si Otto et Anna étaient venus, elle serait ennuyeuse; or, si Waldo était venu aussi, elle serait animée; ... »⁴⁶

Aux yeux de Lewis, ce genre de suite est une propriété « logique » des contrefactuels, dont il rend compte dans sa sémantique. À l’opposé, puisque la distinction entre « logique » ou « sémantique » et « pragmatique » est minorée, sinon complètement dépourvue de sens dans l’analyse proposée ici, il y a *deux* manières d’analyser cet exemple. L’une considère que la suite de contrefactuels a trait à une suite de contextes, chacun permettant d’évaluer un nombre restreint des contrefactuels, l’autre considère qu’elle renvoie à une structure au sein d’un seul contexte. Selon la première analyse, il s’agit d’un phénomène relevant de la « pragmatique », selon la dernière, il d’agit de quelque chose de « sémantique ».

Pour commencer avec l’analyse pragmatique, on pourrait analyser le contrefactuel, $A_1 \Rightarrow B$, dans un cadre \mathcal{C} ⁴⁷, avec l’algèbre engendrée par A_1 , B , et R avec $B \equiv A_1 \vee R$, et l’ensemble co-indépendant $\{A_1, R\}$ ⁴⁸. Pour reprendre la notation introduite dans §4.2.1, on a $\vdash_{\mathcal{C}} A_1 \Rightarrow B$ (le contrefactuel est validé dans le cadre \mathcal{C}).

$A_1 \& A_2 \Rightarrow \text{non } B$ ne trouve pas d’analyse dans \mathcal{C} puisque A_2 y est absent. Lorsque l’on considère le contrefactuel « si Otto était venu, la soirée serait animée » (dans \mathcal{C}), l’assistance de Anna est *hors sujet*. Pourtant, il y a des distensions de \mathcal{C} où ce contrefactuel pourrait être évalué. Un exemple est le cadre consistant en l’algèbre engendrée par A_1 , A_2 , B et R avec $B = (A_1 \wedge \neg A_2) \vee R$, et l’ensemble co-indépendant $\{A_1, A_2, R\}$, que l’on appelle \mathcal{C}' . \mathcal{C}' est une distension de \mathcal{C} ; en outre, $\vdash_{\mathcal{C}'} A_1 \wedge A_2 \Rightarrow \neg B$ et $\vdash_{\mathcal{C}'} A_1 \Rightarrow B$ – les deux premiers contrefactuels dans la suite sont validés dans le cadre \mathcal{C}' .⁴⁹

⁴⁶Lewis (1973b, p10).

⁴⁷Pour alléger la discussion on travaille avec les cadres plutôt que des collections de cadres, ou autrement dit, avec des contextes comportant un seul cadre (§5.1.5).

⁴⁸ R , le « reste », est intuitivement toutes les autres manières dont B arriverait (dont la soirée serait animée); voir §5.1.3. On suppose que R est faux réellement, car, sinon, l’assistance d’Otto n’aurait aucune conséquence sur la soirée, et on frôlerait les problèmes du genre du *verum ad quodlibet*. On fait le même genre de supposition pour les autres cadres considérés ci-dessus.

⁴⁹À strictement parler, on affaiblit un peu l’analyse (**ALP**) pour permettre d’analyser des contrefactuels qui dont l’antécédent est la conjonction de plusieurs éléments de l’ensemble co-indépendant (l’analyse ressemble plutôt à la Satisfaction_r de §A.2 (Définition A.2.1)). Ce détail ne touche pourtant aucunement à l’argument. Par ailleurs, le fait de considérer les ensembles co-indépendant à la négation de leurs éléments (Remarque 5.1.2 de §5.1.2) permet l’emploi d’un même ensemble co-indépendant pour évaluer plusieurs contrefactuels.

Or, $A_1 \& A_2 \& A_3 \Rightarrow B$ ne trouve pas d'analyse dans \mathcal{C}'' puisque A_3 y est *hors sujet*. Alors, on passe à une distension \mathcal{C}' où $\vdash_{\mathcal{C}'} A_1 \wedge A_2 \wedge A_3 \Rightarrow B$, $\vdash_{\mathcal{C}'} A_1 \wedge A_2 \Rightarrow \neg B$, et $\vdash_{\mathcal{C}'} A_1 \Rightarrow B \dots$

La manière « pragmatique » de comprendre cette suite de contrefactuels est comme une montée des cadres (et corrélativement, de contextes) de plus en plus « distendus », permettant l'évaluation d'un ensemble accroissant de contrefactuels de la suite. La manière « sémantique » ou « logique » de comprendre la suite se définit simplement par rapport à cette suite de cadres : il s'agit de prendre le cadre « *limite* » de cette ascension, lequel comporte tous les antécédents et permet d'évaluer tous les contrefactuels de la suite. Ce cadre limite est une structure « *globale* », au sens où il permet d'évaluer tous les contrefactuels (du moins, de cette suite), qui en quelque sens résume tous les cadres individuels, *locaux* dans le sens où ils ne peuvent qu'évaluer un nombre restreint de contrefactuels de la suite.

Dans l'exemple considéré, c'est *grosso modo* cette structure globale que proposent les théoriciens pour analyser la suite de contrefactuels (on a constaté dans §4.2.1 que leurs analyses sont globales). On *retrouve* d'ailleurs les analyses traditionnelles, telle celle de Lewis, en employant la perspective « extensionnelle » sur l'algèbre de ce cadre « limite », consistant à y voir des (petits) mondes possibles (§§4.2.1 et 5.1.1). Dans cette perspective, il y a des « mondes » où les propositions B , A_1 , A_2 , A_3 , ... sont soit vrais soit faux, tels que dans tous les « mondes » où $A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge \dots$, B est vrai (dans les « mondes » où Otto est venu, mais non pas Anna, ni Waldo, ni ..., la soirée est animée), et dans tous les « mondes » où $A_1 \wedge A_2 \wedge \neg A_3 \wedge \dots$, $\neg B$ est vrai (dans les « mondes » où Otto et Anna sont venus, mais non pas Waldo, ni ..., la soirée est ennuyeuse), et ainsi de suite. En définissant une relation de similitude sur ces mondes selon laquelle les « mondes » où $A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge \dots$ sont « plus proches » du monde actuel que ceux où $A_1 \wedge A_2 \wedge \neg A_3 \wedge \dots$, qui sont pour leur part « plus proches » que ceux où $A_1 \wedge A_2 \wedge A_3 \wedge \dots$, et ainsi de suite, on retrouve exactement l'analyse que Lewis propose pour cet exemple⁵⁰. À des détails sur les mondes possibles près, l'analyse en termes d'une relation de similitude, et d'ailleurs les analyses de la tradition en général, est *équivalente* à l'analyse que l'on ferait dans un cadre « *global* », qui est la *limite* des cadres locaux employés qui évaluent des contrefactuels individuels.

Cela éclaire d'un jour nouveau le rapport entre la théorie proposée ici et la tradition. Les théories traditionnelles mobilisent des structures « globales ». Il y

⁵⁰ *Op cit.*, p11 *sq.* Sa formulation de son analyse, en termes de sphères, est équivalent à la formulation en termes d'une relation de similitude (voir son §2.3).

a une parenté entre ces structures et les « grands » cadres que l'on trouve dans l'optique développée ici. Ces cadres « résument » en quelque sorte les cadres locaux et ordinaires qui sont normalement employés pour analyser les contrefactuels individuels (ou des petits ensembles de contrefactuels). Ils sont obtenus à partir des cadres locaux par un « passage à la limite ». En somme, on retrouve les théories traditionnelles comme des cas limites de la théorie proposée ici. Par rapport à la tradition, la théorie plus subtile proposée ici ne perd rien ; elle représente un gain absolu.

5.3.2 Concevoir la logique

Où l'on voit que, non seulement le cadre théorique proposé ci-dessus permet de retrouver les règles logiques proposées par d'autres théories des contrefactuels, mais il offre une vision nouvelle sur ce qu'est une règle logique, et ce qu'est la logique en général. Il s'avère que la théorie des contrefactuels proposée ci-dessus n'est qu'un exemple d'une approche – la logique locale – qui a un domaine d'application plus vaste et important. En particulier, en supportant aussi bien une formulation des règles logiques, que des modèles plus fidèles aux usages effectifs des opérateurs ou des connecteurs (tels les contrefactuels), ce genre d'approche permet de mieux comprendre le sens et la portée des règles logiques, et leur rapport à la réalité effective.

Dans son acception la plus générale, la « logique » a trait à la vérité ou à l'assertion, et à leur transmission. Ainsi conçu, les *faits effectifs* et *observables de manière immédiate* qui sont pertinents à la logique ont affaire à l'apparition effective des contrefactuels dans les discours, les raisonnements et les argumentations qui se produisent réellement par exemple, à leurs propriétés en tant qu'ils y apparaissent et aux avis que les gens expriment relatifs à eux. Cet ensemble de « données » est, à première vue, très *complexe* ; une recherche qui vise à les comprendre à leur propre niveau de complexité sera forcément elle-même difficile et compliquée. À l'opposé, on peut poser des structures ou des systèmes plus *simples*, donc *compréhensibles*, qui sont censés se rapporter d'une manière ou d'une autre aux faits effectifs. Dans la mesure où ces structures sont plus commodes à la pensée humaine, une science qui travaille avec elles seules sera plus facile et faisable.

Le problème des contrefactuels posé ici concerne les faits effectifs à leur égard ; en conséquence la théorie des contrefactuels esquissée est du premier type décrit ci-dessus. Par opposition, les recherches proposées traditionnellement, en se limitant au dit problème « sémantique », concernent plutôt les systèmes et les

structures plus simples et maniables. Telles sont les recherches qui étudient et proposent des « règles logiques » : ces règles, et les systèmes qu'elles constituent, sont relativement simples. Par conséquent, comme l'a été déjà souligné dans §4.3.1, il y a un certain décalage entre les règles simples et claires proposées et la réalité complexe et vague relative aux comportements des contrefactuels en pratique. On dit souvent que les règles, et plus généralement les structures qu'elles impliquent, sont des « *idéalisations* » ou des « *abstractions* » à partir de la réalité effective et complexe, ou autrement des « *approximations* ». Cette distance entre les systèmes logiques proposés précédemment et les faits effectifs avait pour conséquence que ces systèmes portaient peu d'intérêt pour la recherche menée dans cette thèse, laquelle visait les faits eux-mêmes, plutôt que quelconque « *idéalisation* ». Les théories des contrefactuels proposées dans la tradition, et les systèmes et règles logiques qu'ils impliquent, *sont nettement dépassées* par une théorie qui rend compte des complexités réelles elles-mêmes. En revanche, si ces systèmes logiques sont des « *abstractions* » ou des « *approximations* » relatives aux faits, il devrait être possible de les retrouver à partir d'une théorie complexe fidèle aux faits en effectuant les abstractions appropriées. À partir du simple, il n'est pas possible de regagner le complexe ; munie de la mécanique newtonienne, seulement un saut significatif mène à la relativité générale. Par contraste, le passage inverse, du complexe au simple, est généralement plus facile, consistant à dépouiller les complexités appropriées ; du point de vue de la relativité générale, la mécanique newtonienne est immédiatement reproduit comme cas spécial.

Ainsi, on peut considérer les règles logiques qui ont été proposées pour les contrefactuels comme un défi pour le cadre théorique proposé ici : celui-ci devrait permettre de les retrouver comme des « cas spéciaux » correspondant à des *simplifications appropriées*. Ce défi est tant *conceptuel* que *technique* : à quoi, dans un cadre théorique qui s'occupe des contextes et des transitions entre eux, pourrait référer une règle logique ? En réponse, ce cadre offre une distinction entre *trois statuts* différents que pourrait avoir une règle logique, voire même *trois sens* de ce que c'est un système logique, et par conséquent trois manières de comprendre les systèmes logiques proposés dans le passé. En effet, une règle ou un système logique pourrait se rapporter à une des trois « structures logiques » suivantes — la structure logique *locale* en vigueur à un instant donné — la structure des opérations prenant des structures locales aux nouvelles structures locales — la structure « globale » qui est « sous-jacente » à, ou en quelque sens « résume » une collection de structures locales. On expose une par une les idées des trois « champs » d'une règle logique, en constatant dans chaque cas l'abs-

traction requise pour obtenir des règles logiques « simples » à partir du modèle complexe proposé. Les considérations techniques ainsi que la comparaison avec les règles logiques proposées dans la tradition sont reportées à l'Annexe A.

La clef de voûte de la conception et la théorie des contrefactuels développées ici est le contexte, en tant que structure *logique*, *vague* et *locale*. Une règle logique pourrait être comprise comme référant à une propriété d'une telle structure ; par exemple la règle avec prémisses $A \Rightarrow C$ et $A \leftrightarrow B$ et conclusion $B \Rightarrow C$ ((**ELG**) de §A.1) correspond au fait que si le contexte valide le premier contrefactuel et l'équivalence y vaut, alors il valide le dernier contrefactuel. Les règles logiques sont précises ; par conséquent, le genre de structure qu'elles décrivent ne comporte pas de vague. L'idéalisation qui produit la structure décrite par les règles logiques consistera à remplacer les contextes vagues par des structures précises, en l'occurrence par les cadres (§5.1.2, Définition 5.1.8). En général, selon cette interprétation des règles logiques, elles ont trait à la structure des *cadres locaux*. La discussion technique de cette conception des règles logiques se trouve dans §A.2.

Il y a des transitions entre des contextes différents ; une règle logique pourrait avoir affaire aux propriétés de ces transitions. Par exemple, la règle (**ELG**) mentionnée ci-dessus correspondrait au fait que, si l'on a d'un cadre où le premier contrefactuel est valide, et un autre cadre où l'équivalence est valide, on pourrait passer dans un cadre où le contrefactuel dans la conclusion est valide. Cette acception des règles représente une simplification de la situation effective : normalement, la dynamique de contextes consiste en des glissements complexes, flous et non strictement réglementés, alors que selon cette conception les règles logiques décrivent des *opérations précises et bien définies* sur cadres, prenant un ou plusieurs à un autre. Une règle logique, ainsi conçue, est une *règle de transition entre cadres locaux*. §A.3 contient une discussion technique sur cette conception de règle.

Les contrefactuels sont évalués dans les contextes ou (d'après la simplification pour permettre des règles précises) des cadres locaux, dont chacun ne permet d'évaluer qu'un nombre relativement restreint de contrefactuels. On pourrait néanmoins se demander s'il existe, pour une collection de cadres locaux donnés, une structure capable à évaluer *tous* les contrefactuels qui sont évalués dans *au moins* un des cadres locaux et qui est *généralement* d'accord avec les cadres locaux sur leurs évaluations. La suite de contrefactuels considérée dans §5.3.1 fournit un exemple : le cadre « global » qui est la limite de la suite de cadres locaux, dont chacun évalue certains contrefactuels mais non pas tous, évalue

tous les contrefactuels, et les accorde la même valeur de vérité que les cadres locaux individuels. Une telle structure globale *résume* en quelque sorte les cadres locaux. Une règle logique pourrait, finalement, être comprise comme renvoyant à une propriété d'une telle structure globale. La discussion des complexités de trouver de telles structures globales, et les règles logiques qui s'y appliquent est reportée à §A.4.

Le cadre conceptuel proposé n'offre donc pas seulement la possibilité de retrouver les résultats, les règles et les techniques développés par les logiciens : elle offre une conception renouvelée du sens et du champ d'application de ces règles, ces résultats et ces techniques. En un mot, elle offre une conception renouvelée de *ce qu'est la logique*.

D'après cette conception, la logique s'occupe des *cadres locaux* – des structures logiques qui sont locales au sens où ils ne comportent pas « tous » les énoncés –, de leurs éventuelles structures internes, des transitions et des opérations sur eux, et des structures plus « globales » qui « résument » des collections de cadres locaux. Cette manière de traiter les questions normalement qualifiées de logiques, et la collection d'outils et d'acquis de cette approche, s'appelle la *logique locale*.

La logique locale est donc plutôt *une manière de concevoir et de faire la logique qu'un système particulier*. La considération de contrefactuels présentée dans ce travail pourrait être traitée comme une étude de cas pour une technique beaucoup plus général. La logique locale pourrait être appliquée pour étudier d'autres systèmes logiques et d'autres constructions logiques. En jouant sur les définitions de cadre local, et sur les transitions et les opérations permises sur eux, on pourrait retrouver non seulement la logique propositionnelle, mais la logique du premier ordre ou les logiques modaux⁵¹ ; on pourrait étudier non seulement les contrefactuels, mais les opérateurs de croyance ou de connaissance, et les opérations dynamiques sur eux⁵².

La logique locale est pertinente aussi bien pour des questions dites « syntaxiques » que celles que l'on qualifie habituellement de « sémantiques ». D'un côté, la conception des règles logiques comme des contraintes sur les opérations de transition entre les cadres convient à la pensée des preuves : en modélisant les séquents qui apparaissent à des étapes des preuves par des cadres particuliers, et les règles par des opérations sur les cadres, la perspective de la logique

⁵¹Par exemple, il s'agit de remplacer l'algèbre de Boole comme algèbre de base pour le cadre par l'algèbre cylindrique (Henkin et al., 1985a,b; Monk, 2000) ou l'algèbre modale (Blackburn et al., 2001, Ch. 5)

⁵²Il y a, par exemple, un rapport étroit et de longue histoire entre les contrefactuels et la révision de croyances ; voir les références mentionnées dans la note 4.

locale fournit des modèles intéressants pour la théorie de la démonstration⁵³. De l'autre côté, les questions sémantiques se bifurquent dans l'optique locale : tantôt on traite la structure logique d'un cadre local individuel, tantôt on traite la structure plus « globale » qui « résume » une collection de cadres locaux. Les techniques et les résultats des théories sémantiques établies peuvent être appliquées, avec plus ou moins de pertinence, aussi bien au niveau local que global. En effet, dans l'emploi habituel de ces théories, il y a trop souvent de confusion sur le niveau qu'ils traitent. Par exemple, la confusion des théories traditionnelles de contrefactuels qui a été soulignée dans §4.3.1 est sans doute liée à une telle ambiguïté : les théories sémantiques proposées peuvent être considérées comme « locales » dans la mesure où ils permettent un apport pragmatique qui pourrait changer d'un contexte à l'autre, mais elles sont « globales » dans la mesure où elles insistent pour évaluer tous les contrefactuels du langage (§4.2.1). La perspective de la logique locale permet de mettre au clair la situation, et par conséquent d'éviter ces confusions nuisibles.

Enfin, la logique locale s'applique non seulement aux questionnements relatifs aux règles logiques strictes et précis, mais également à ceux qui concernent les propriétés complexes et flous de discours, de raisonnements, d'argumentations réelles. La théorie proposée dans §5.1, comme on l'a suffisamment souligné, s'inscrit dans ce genre de problématique, dans la mesure où elle a trait au comportement effectif de contrefactuels. Cette capacité de supporter et d'exprimer aussi bien les recherches relatives aux systèmes simples et précis, tels les systèmes logiques proposés habituellement, que celles concernant les phénomènes complexes et vagues constitue un *atout majeur* de la logique locale. Tout simplement, le fait de pouvoir exprimer les deux genres d'enquêtes, et leurs résultats, dans une forme unique permet d'approfondir la compréhension du rapport entre eux. Ce qui est important philosophiquement. Car, comme l'a été discuté au début de ce paragraphe, les systèmes simples et précis, et les règles logiques qu'ils comprennent, sont le plus souvent considérées être des « idéalisation » ou des « approximations » des faits effectifs dont les théories complexes rendent compte de manière plus exacte. Quel que soit le sens que l'on donne à cette « idéalisation », une compréhension du rapport entre un système de règles logiques et une théorie complexe ayant le même objet aura une importance notable. D'un côté, si « l'idéalisation » est *normative*, si une règle logique est une contrainte sur ce que l'on *devrait* faire ou dire, alors en comprenant le rapport

⁵³Par exemple, une sémantique de preuves pour la logique classique pourrait être développée dans cette optique; voir Hill (2005), *Une sémantique de preuves pour la logique classique*, Mémoire de Masters 2, Logique Mathématique et Informatique Théorique, Université Paris 7.

entre la règle logique et la théorie complexe, laquelle est entièrement *descriptive*, on comprend le rapport entre le *normatif* et le *descriptif* quant à cet objet, entre le *ce que l'on affirme* et le *ce que l'on est permis d'affirmer* avec tel opérateur (par exemple). De l'autre côté, si « l'idéalisation » est une simplification théorique visant à s'approcher de la réalité effective, toujours dans un but descriptif, alors une compréhension du rapport avec la théorie complexe qui est plus proche aux faits effectifs fournit une compréhension de la fécondité, de l'exactitude et du domaine d'utilité de cette « idéalisation ». Dans les deux cas, il s'agit des aspects fondamentaux pour comprendre la teneur réelle des règles logiques.

Le cadre conceptuel introduit dans §5.1 peut non seulement rendre compte des théories traditionnelles des contrefactuels, dans la mesure où elle peut reproduire les règles logiques qu'elles proposent ; également, et de plus important, il offre de nouvelles perspectives sur ce qu'est une règle logique, et une nouvelle conception de la logique en générale. Cette conception s'applique loin au-delà du cas des contrefactuels, se montrant propice relativement aux questions « sémantiques » aussi bien que « syntaxiques », supportant des considérations ayant trait aux systèmes précis et idéalisés ainsi que celles qui s'occupent des faits effectifs complexes. Par suite, elle permet de comprendre les rapports entre ces différents domaines et ces différentes recherches, ce qui ne peut qu'approfondir la compréhension des rapports entre eux. En faisant l'économie des considérations traditionnelles relatives aux règles logiques dans le développement de la théorie, on a su proposer une théorie et une perspective qui les dépassent de loin, mais qui, en retour, permettent de les concevoir sous un jour nouveau. La théorie proposée des contrefactuels n'est qu'un début. Elle a ouvert un champ, celui de la logique locale. Désormais, comme à la suite de tout pas en avant, beaucoup reste à faire.

Conclusions et Projections

On a vu qu'entre la fiction et le contrefactuel, il y a une parenté de questionnement. Dit grossièrement, une question commune et importante se pose à leur égard : celle de leur rapport à la réalité. D'un côté, les vérités fictionnelles dépendent des vérités réelles, mais également certaines vérités réelles peuvent apparemment s'éclairer à la lumière de la lecture des textes fictionnels ; les croyances, les préjugés et les attitudes ordinaires du lecteur jouent un rôle important dans sa lecture des textes de fiction, qui peuvent exercer en retour des effets sur ces attitudes mentales à l'égard de la réalité. De l'autre côté, la vérité d'un contrefactuel s'appuie le plus souvent sur des faits réels, mais inversement le fait pour un contrefactuel d'être vrai a des conséquences sur ce qui peut arriver en réalité.

Cette communauté de questionnement est accompagnée d'une communauté de réponses possibles. En effet, deux approches générales de ces questions s'opposent. D'un côté, l'approche traditionnelle tend à poser d'abord un « champ fixe d'états », et ensuite une structure supplémentaire, souvent de l'ordre d'une hiérarchie ou une priorité sur ces états, qui est censée permettre de penser la fiction ou les contrefactuels en termes de ces états fixes. La stratégie « mimétique » (§1.2.1) consiste à poser deux choses opposées – l'une est la réalité, l'autre la fiction –, pour ensuite remarquer un rapport complexe entre elles. Les approches (conceptuelles) du mental (§§1.3 et 1.4) admettent une collection fixe d'énoncés envers lesquels le sujet peut avoir des attitudes mentales, pour ensuite accorder des statuts (ou des degrés) de croyance ou de *faire comme si* à ces énoncés. Enfin, l'approche la plus répandue des contrefactuels (§4.2.1) consiste à poser une collection fixe de mondes possibles, pour ensuite y ajouter une relation de similitude sur ces mondes possibles permettant d'évaluer les contrefactuels.

Pourtant, comme on l'a montré dans les Chapitres 1 et 4, ce genre de démarche ne parvient pas à rendre compte premièrement de la dynamique de la structure supplémentaire qu'elle propose, et deuxièmement de certaines propriétés particulières de la fiction ou des contrefactuels en termes de cette

structure supplémentaire, dont notamment le phénomène du *hors sujet*. Non seulement les approches mentales de la fiction ont peine à rendre compte de la complexité des rapports entre les attitudes mentales envers la fiction et la réalité à des moments particuliers, mais elles échouent largement à éclairer la dynamique de ces rapports (§§1.4 et 1.5). De manière semblable, les théories traditionnelles des contrefactuels font fi aussi bien des propriétés fines des contrefactuels dans des situations particulières, tels les phénomènes de pertinence de l'antécédent par rapport au conséquent ou le *hors sujet* de certains énoncés, que de la dynamique de la relation de similitude qu'elles supposent de l'autre.

De l'autre côté, on a développé dans les Chapitres 2, 3 et 5 une approche opposée, qui prend comme point de départ les structures *locales*, figurant seulement les aspects mentaux, linguistiques ou autres qui sont en jeu dans des situations particulières : c'est-à-dire, dans le cas de la psychologie de la fiction, les structures instantanées de la conscience (§§2.1–2.3), et, dans l'analyse des contrefactuels, les contextes qui permettent de les évaluer (§5.1). Ce faisant, l'approche rend compte naturellement du *hors sujet* en tant qu'absence par rapport à la structure locale (§2.2 et §5.2.1). Ensuite, on s'est posé immédiatement les questions *externes* à ces structures locales : celles qui ont trait d'une part aux relations, notamment dynamiques, entre les structures locales différentes, et d'autre part aux propriétés des collections ou des successions de ces structures. En premier lieu, on a pu, en s'appuyant sur le vague de ces structures, entamer une réponse aussi bien à la question de la dynamique des structures instantanées de la conscience (§2.3) qu'à celle de la dynamique des contextes (§5.1.5). Ce qui a permis d'esquisser des résultats approfondis relatifs à la lecture de la fiction et aux contrefactuels (§§3.1–3.2 et §5.2.3). En second lieu, on a pu retrouver, en tant que propriétés des collections ou des successions de structures locales, les notions, voire les « états fixes », qui soutiennent les analyses traditionnelles de la fiction et des contrefactuels, à savoir les États Mentaux tels la croyance et le *faire comme si* d'un côté, et les mondes possibles munis d'une relation de similitude de l'autre (respectivement, §§2.5 et 3.3, et §§5.3.1 et A.4.2).

Cette opposition est de prime abord *méthodologique*. Il s'agit de deux stratégies ou approches générales par rapport au même sujet. L'une le prend d'en haut, pour ainsi dire, en posant un « champs d'états » fixe, immuable, un *sol*, et en espérant rendre compte des phénomènes avec une structure supplémentaire *sur* ce sol, qui organise ses « états ». L'autre le prend d'en bas, si l'on peut dire, en cherchant le sujet d'abord dans les localités ou les *sites* appropriés, pour réunir après coup les données des sites différents. La première approche est *globale*, la deuxième approche est *locale*.

Néanmoins, dans la mesure où toute enquête, qu'elle soit métaphysique, ontologique, épistémologique ou autre requiert voire suppose une méthodologie, la question se pose des conséquences de cette opposition sur les plans métaphysique, ontologique ou épistémologique. Les conséquences sont tout probablement délicates. En particulier, l'approche locale *ne consiste pas à nier* le *sol* et les concepts posés par l'approche globale – elle parvient d'ailleurs à les reproduire – mais à *nier seulement* que ce *sol* et ces concepts permettent d'aborder les questions posées (sans ajouts supplémentaires). Si la métaphysique ou l'ontologie se règle sur l'*emploi* que l'on fait de leurs concepts, ce qui est d'ailleurs le cas chez des philosophes aussi différents que Wittgenstein, Saint Thomas, Descartes, Kant et Quine⁵⁴, alors il s'agit de nier que le *sol* et les concepts posés par ces approches globales soient *bons*. Comme on verra prochainement, l'approche locale peut en venir finalement à proposer un autre *sol* et d'autres concepts globaux, qui « résument » en quelque sorte toutes les structures locales (voir également §A.4.2). La question des implications métaphysiques, ontologiques, ou épistémologiques de l'approche locale, et de son attitude à l'égard de la question du *sol*, et plus généralement la question des enjeux de ce conflit d'approches sur d'autres plans philosophiques indique une éventuelle direction de recherches qui prolonge les considérations présentées ici.

Par ailleurs, les travaux effectués ici pourraient se poursuivre dans une direction différente. Si l'enjeu est méthodologique, on s'attend à ce que les conclusions des présentes recherches s'appliquent au-delà des questions de la fiction et des contrefactuels. Et, en effet, la problématique formulée à l'égard de la fiction et des contrefactuels trouvent un écho dans d'autres domaines, de telle sorte que l'opposition des approches présentée ici s'y reproduit. Pour terminer, il s'agit d'évoquer certains de ces domaines, et d'esquisser le programme qui consiste à étendre l'approche locale à ces questions.

Il paraît possible de rassembler les phénomènes considérés dans cette thèse sous le nom des phénomènes de la *potentialité* au sens large. L'entrée en jeu des faits réels dans la lecture de la fiction, ainsi que l'empiètement des aspects de la lecture de la fiction sur la conception de la réalité, ne sont-elles pas des projections au sens de Goodman, et donc de l'ordre de la potentialité⁵⁵ ? Également, n'y a-t-il pas l'idée qu'un contrefactuel tel « si l'allumette était frottée, elle ne s'allumerait pas » exprime une « poussée vers » ce qu'elle allume dans le cas

⁵⁴La différence entre ces philosophes consiste en le type d'emploi qui est privilégié : pour Wittgenstein, mais également pour Saint Thomas, c'est l'emploi dans la vie quotidienne (comme le remarque Pouivet (2006)), pour Descartes et Kant, c'est un emploi apodictique, pour Quine, c'est l'emploi dans la science.

⁵⁵Goodman (1954, notamment Ch 2–4).

de frottement, une plus grande « potentialité » de cette éventualité que de son inverse ?

Ce rapprochement ne repose pas que sur une intuition : les structures que l'on a repérées auprès de la fiction et des contrefactuels, ainsi que les questionnements que l'on a formulés à leur égard, trouvent un écho du côté de la potentialité. Comme la fiction et le contrefactuel, la potentialité s'oppose à la réalité (au sens étroit) ou si l'on préfère l'actualité. Or, comme dans les cas de la fiction et du contrefactuel, elle entretient un rapport complexe avec cette réalité : non seulement il y a dépendance de ce qui est potentiel relativement à ce qui est réel, mais inversement les potentialités ont des conséquences importantes par rapport à la réalité (ou l'actualité, si l'on préfère). Si bien que l'on peut esquisser deux approches quant à la question de la potentialité : une approche globale et une approche locale. En (re)présentant les grandes lignes de ces deux approches, on esquisse non seulement deux approches de la potentialité en général mais on fait apparaître éventuellement l'opposition discutée ci-dessus sous un nouveau jour. On introduit les deux approches par le moyen d'un exemple, en l'occurrence celui du dialogue suivant (ayant lieu dans un endroit peu venteux).

A — Si je frottais cette allumette, elle s'allumerait.

B — Pas nécessairement : ce ne serait pas le cas s'il y avait un soudain coup de vent par exemple.

A — Certes, s'il y avait un soudain coup de vent et je frottais l'allumette, elle ne s'allumerait pas. Pourtant, bien que possible, l'éventualité d'un soudain coup de vent n'a que peu de chance de se produire. Ainsi, je persiste à dire que si je la frottais, elle s'allumerait.

Voici une manière de comprendre ce dialogue. L'interlocuteur B, en relevant la question du vent, a bien identifié un événement possible. Au cas où cette possibilité s'actualisait, elle invaliderait le conséquent du contrefactuel affirmé. Pour se défendre contre l'atteinte portée par cette possibilité à la vérité de sa première affirmation, il fallait que A rend compte de sa potentialité ou probabilité de se produire ; ce faisant, A a pu la rejeter comme possible mais assez peu probable.

En voici une autre. L'interlocuteur B, en mettant en jeu le vent, a effectivement opéré un glissement du sujet. A aurait pu bien se montrer impatient envers ce changement du sujet ; cependant, il a choisi de suivre B et de poursuivre la discussion « dans ses termes ». Mais A l'a suivi seulement jusqu'à un certain point : il admet que le conséquent serait démenti en cas du vent, mais il récuse la pertinence de la supposition du vent – non pas au titre d'un glissement

conversationnel, mais plutôt, dans le cadre du nouveau contexte, sous le nom d'une trop faible potentialité.

En un mot, selon la première approche, le dialogue se déroule dans un seul et même contexte ; selon la deuxième, il y en a au moins deux. On reconnaît la première approche aux partisans des approches globales⁵⁶ ; la deuxième se rapproche du côté de l'approche locale.

On peut comprendre les différences entre ces deux styles d'approches sur cet exemple simple. Si l'on comprend le dialogue de la première manière, on est souvent peu enclin à se permettre d'être « coupé » par son interlocuteur : on ne permet pas qu'il mette en jeu un énoncé dont on ne sache pas (*déjà*) le statut. On fait en sorte, donc, qu'il n'arrivera jamais de surprises⁵⁷. C'est-à-dire, qu'il n'arrivera jamais de surprises, *tout en restant dans ce même contexte*. Cette minutie tatillonne, combinée avec l'unicité du contexte, a pour effet de gonfler le contexte de telle manière qu'il « contient » et pourrait « évaluer » tout ce qui est « admissible », dans le sens le plus grand de ce terme. Alors, tout énoncé du langage devrait y trouver une interprétation, et notamment tout contrefactuel. Ce contexte est l'enseigne de l'approche globale.

En revanche, la crainte d'être surpassé par l'interlocuteur, la crainte qu'il nous laisse muet, ne trouve pas de raison dans la deuxième approche au dialogue. Puisque, où qu'il aille, nous pouvons toujours suivre en nous plaçant dans le contexte qu'il pose de nouveau. Plus : ce mouvement est la condition même de la communication ; il est une des formes fondamentales de la coopération sur laquelle toute communication se repose. Aucune nécessité donc de nous protéger en hypostasiant un contexte immense ; la confiance et l'intelligence de suivre notre interlocuteur autant qu'il faudra, et d'employer les outils qu'il a mis lui-même à notre disposition suffiront. S'esquisse ainsi la figure de l'approche locale.

Mais quelle conséquence a cette différence de contextes pour la potentialité ? Celle-ci : si l'on admet que la discussion du vent et du frottement de l'allumette concerne d'autres « possibilités », lesquelles sont plus ou moins « potentielles », alors le contexte impliqué dans l'approche globale comporte (*déjà*) *toutes les* possibilités, dans quelque sens absolu, alors que le contexte *local* employé dans l'approche opposée ne contient que les possibilités « en jeu » ou « envisagées » à ce moment. L'approche globale s'appuie donc sur un champ ou *sol fixe, grand* et *global* de « possibilités », lequel délimite la totalité de ce qui pourrait entrer en jeu, où « le degré de potentialité » consiste en la manière dont ces possibilités

⁵⁶ Cf. Lewis (1973b, p10), par exemple, discuté dans §5.3.1.

⁵⁷ Cf. Wittgenstein, (1922, §6.1251 ; 1969, §71).

sont organisées ou ordonnées. Par contraste, dans l'approche locale, il ne s'agit pas, du moins de prime abord, de « toutes les possibilités » au sens absolu de l'approche globale, mais plutôt des « possibilités » qui sont *en jeu* ou *en vigueur*, voire même *actives*, et leur « degrés de potentialité » sont simplement ceux qu'ils ont en tant qu'ils figurent dans ce contexte. Entre contextes, de même qu'entre moments temporels, les possibilités qui sont pertinentes, effectives ou actives changent, ainsi que leurs « degrés de potentialité ». Pour retrouver quelque chose de l'ordre des « possibilités globales » mise en avant par l'approche globale, il faut considérer toute la gamme de contextes différents, les rapports entre eux et leurs propriétés générales, et leur prendre la « moyenne », pour ainsi dire.

L'approche globale de la question de la potentialité commence donc par la factoriser : d'un côté, il y a un *sol fixe* et *sempiternel*, capable d'accueillir tout ce qui peut entrer en scène (en un sens large de « peut ») ; de l'autre côté, il y a une structure supplémentaire *sur* ce sol, qui organise les « points » ou les « états » de ce sol selon qu'ils soient plus ou moins potentiels. L'étape la plus importante et d'ailleurs la plus difficile dans cette approche est la première : de trouver une *factorisation féconde*, de poser le *bon sol*. Les aléas de sa première étape rend la démarche fragile. En revanche, la voie empruntée par l'approche locale est plus sûre, dans la mesure où son premier geste est modeste : il consiste à accepter et caractériser seulement ce qui est donné immédiatement dans le *site* approprié. Or, ce n'est pas à dire qu'elle se limite ou s'enferme dans ces sites ; au contraire, tout le travail commence après la caractérisation des sites individuels, lorsque l'on en vient à aborder la question de la dynamique des potentialités (ou, si l'on préfère, la dynamique des « degrés de potentialité ») et à rassembler tous les sites dans une structure unique et globale, un *sol* non pas posé mais trouvé. Si l'on voulait résumer l'opposition dans une phrase elle serait la suivante : ce que l'approche globale devine, l'approche locale le découvre.

L'identification de la potentialité, au sens large, comme comportant une problématique semblable à celles de la fiction et des contrefactuels, et par conséquent comme susceptible à deux styles d'approche importe en raison de la multitude de domaines philosophiques qui comportent des phénomènes se rattachant à quelque chose de l'ordre de la potentialité. D'ailleurs, les théories proposés dans ces domaines s'avèrent être des exemples d'approches globales ou dans une moindre mesure locales. La continuation naturelle du présent travail consistera à appliquer l'approche locale dans certains ou tous ces domaines, et surtout ceux qui ne connaissent pas cette approche.

Tout d'abord, la potentialité elle-même pourrait être étendue dans un sens « métaphysique », en tant qu'une sorte de « propension » ; sous cette dimension,

elle se rapporte à d'autres questions telles celles de la causalité. De manière générale, les théories traditionnelles de la potentialité, de la propension, ou de la causalité empruntent l'approche globale. Certaines admettent comme *sol* un champ de substances et pensent la potentialité comme une propriété spécifique des substances, c'est-à-dire comme relevant de ce que l'on *structure supplémentaire* aux substances⁵⁸. D'autres conçoivent les propensions, et ensuite la causalité, en termes de mesures de probabilité (*structure supplémentaire*) sur un espace d'événements ou situations (*sol*)⁵⁹. D'autres encore pensent la causalité et la probabilité (objective) en termes de contrefactuels, et par suite en termes de similitude (*structure supplémentaire*) entre mondes possibles (*sol*)⁶⁰.

Par ailleurs, certaines questions qui pourraient être qualifiées de plus ou moins épistémologiques ou doxastiques ont des éléments qui peuvent être rapprochés de la potentialité en sens général. D'un côté, on a déjà rencontré dans §1.3.2 l'idée qu'une croyance est de l'ordre d'une « propension » à agir, représentée formellement comme une mesure de probabilité (*structure supplémentaire*) sur un ensemble de « mondes possibles » (*sol*)⁶¹. En outre, la question des changements des croyances dans des éventualités différentes porte l'enseigne d'une question reliée à la potentialité, comme en témoigne d'ailleurs le rapprochement répandu mais complexe avec les contrefactuels⁶². D'un autre côté, la question de l'induction est, comme l'a bien souligné Goodman par exemple⁶³, une question de potentialité (en l'occurrence, de la conclusion inductive sur la base des observations). En effet, même son approche du problème est globale dans la mesure où elle consiste à doter les éléments d'un ensemble fixe et stable composé de tous les prédicats d'un langage (le *sol*) avec des indices « enracinements épistémiques » (*structure supplémentaire*) apparemment reliés à leur relative *correction* pour emploi dans des inductions⁶⁴.

Enfin, on pourrait voir dans certaines questions de la philosophie du langage l'empreinte de la potentialité. Le sens d'un terme linguistique entre dans un rapport étroit avec les usages que l'on en fait, que l'on pourrait en faire, et que l'on en ferait ; autrement, les emplois passés du terme sont reliés à ses emplois futurs⁶⁵. Considérées sous cet angle, un grand nombre de théories du

⁵⁸Aristote (1991).

⁵⁹Popper (1990).

⁶⁰Lewis (1973a) et (1986c, pp94-7).

⁶¹Ramsey (2003); Savage (1954).

⁶²Voir sur ce point la note 4 de §4.1.2. En conséquence de la parenté entre la révision des croyances et les contrefactuels, une partie de la discussion précédente de ceux-ci s'applique également à celle-là.

⁶³Goodman (1954, Ch. 2).

⁶⁴*Op. cit.*, Ch. IV, §§ 3 – 5.

⁶⁵Le rapport avec la potentialité est particulièrement clair dans le rapprochement de la

langage s'avèrent emprunter l'approche globale. En effet, chacune pose une collection fixe et stable d'« entités du sens » (*sol*) – qu'elles soient des propositions, des énoncés, ou même des expressions individuelles –, pour ensuite ajouter des appareils plus ou moins techniques (*structure supplémentaire*) – qu'elles soient des « coordonnées contextuelles », des « conditions et règles illocutionnaires », des « maximes de conversation » ou autres⁶⁶ – qui servent à ranger les énonciations possibles selon leurs effets illocutionnaires, leurs pertinences, voire leurs différentes sortes de « potentialités » au sens large. Dans cette optique, les longs débats sur la frontière entre la sémantique, entendue comme le domaine du sens, et la pragmatique, en tant que domaine d'usage et d'énonciations effectuées en pratique, se ramènent à la question de la *factorisation correcte* du langage « effectif » en une partie sémantique (*sol*) et partie une non sémantique (*structure supplémentaire*)⁶⁷. D'un autre côté, le jeu de langage wittgensteinien⁶⁸ ressemble, par sa localité, aux *sites* qui font la pierre angulaire de l'approche locale, encore qu'il ne s'attarde que très peu sur des questions *externes* relatives à ses jeux de langage, relatives aux rapports entre eux, ou à leur enchaînement dans la dynamique.

Un programme des recherches s'esquisse, qui consiste à étendre et examiner, voire entreprendre une approche locale relative à certaines de ces questions. En partant d'un phénomène aussi spécifique que l'est la fiction, on « projette » loin au-delà. Comment est-ce possible ? Peut-être parce qu'elle entre dans ce rapport difficile et subtil avec la réalité (pré-)établie. Peut-être parce que la fiction, comme l'être qui change, comme la connaissance qui progresse, comme le langage qui évolue, est sur la limite avancée du monde en mouvement. Peut-être parce qu'elle est, entre autres, sur la crête de la vague qui est l'expérience humaine. Si l'approche globale commence par dresser la carte de l'océan, l'approche locale cherche à saisir la vague en tant que telle, dans chacun de ses moments.

Derrière les visages les moins offensifs attendent les défis les plus difficiles. Sans doute la fiction en est un pour la philosophie. En relevant le défi, nous espérons du moins en avoir appris.

question de l'induction que Kripke propose dans sa lecture de Wittgenstein (Kripke, 1982, Ch. 2 et pp58-64).

⁶⁶Lewis (1970, §III; 1979c); Searle (1969); Grice (1989, par exemple Ch 2).

⁶⁷Il y a par ailleurs des philosophes qui posent un sol fixe des énoncés ou des propositions sans employer d'appareil pragmatique pour rejoindre le langage parlé, en préférant plutôt un rapport direct quoique approximatif. Voir notamment Davidson (1984, Ch 9-12), et surtout p137 pour une formulation très claire de la méthodologie globaliste.

⁶⁸Wittgenstein (1953).

Annexe A

La logique locale : application à la logique non-monotone

Comme on l’a vu dans §4.3.1, il existe plusieurs raisons pour se méfier de la distinction stricte et fixe posée par les théories courantes des contrefactuels entre leur côté « sémantique fg et leur côté « pragmatique » : d’ailleurs, l’analyse proposée ne suppose d’ailleurs aucune telle distinction stricte. Or, pourrait-on vouloir objecter, il y a un certain nombre de règles ou « schéma » d’inférence qui sont souvent admises comme appliquant aux contrefactuels : quels statuts ont-elles dans l’optique proposée ci-dessus ? Dans §5.3.2, on a esquissé les grandes lignes d’une réponse à cette question, qui prend la forme d’une conception de la logique que l’on a baptisé la *logique locale*. Il s’agit dans cette Annexe de présenter quelques résultats et observations techniques pour donner une idée plus approfondie des principaux moments de la logique locale.

Pour dire grossièrement, une règle logique permet de lier un certain nombre d’énoncés, appelés « prémisses » avec un autre énoncé, appelé « conclusion », l’idée étant que la conclusion est d’une certaine certaine manière « permise » par les prémisses. Traditionnellement, deux manières de concevoir cette « permission » : D’un côté, selon la manière « sémantique », une règle exprime une *certaine propriété structurelle* du rapport entre le langage et « les états d’affaires » au sens le plus général, à savoir la propriété que, si toutes les prémisses sont vraies dans un état d’affaires, alors la conclusion l’est aussi. De l’autre côté, selon la manière que l’on peut qualifier d’« inférentielle », les règles ca-

ractérisent les affirmations auxquelles est autorisé celui qui admet un certain ensemble d'énoncés (connus ou supposés), à savoir, la règle autorise quelqu'un à admettre la conclusion une fois qu'il a admis les prémisses.

Dans une optique locale, figurant des cadres logiques entrant dans des rapports différents, la manière « sémantique » de comprendre les règles logiques se bifurque, dans la mesure où elle pourrait être comprise *localement* ainsi que *globalement*. De telle sorte qu'il y a les trois manières suivantes de comprendre une règle logique dans cette optique, que l'on a mentionnées dans §5.3.2.

Sémantique locale les règles expriment des propriétés des structures linguistiques qui sont en vigueur à des moments particuliers d'une conversation, d'une argumentation, d'un discours. Pour employer les termes du Chapitre 5, les règles sont comprises comme correspondant aux propriétés (exigées ou supposées) des *cadres locaux*. Elles s'appliquent *dans* les cadres. Dans la mesure où le cadre caractérise, par sa structure et par les énoncés qui y sont tenus pour vrais, un rapport entre le langage – ou plus précisément un fragment de langage – et les états d'affaires, il propose une sémantique ; dans la mesure où le cadre est locale, cette sémantique est *locale*.

Sémantique globale les règles peuvent caractériser des propriétés du rapport entre le langage et le monde qui régissent non seulement chaque cadre logique en vigueur à chaque instant du discours, mais également le discours ou la conversation en entier. Elles ne s'appliquent pas seulement *dans* les cadres logiques, mais les *transcendent* d'une certaine façon. En s'intéressant à la sémantique en ce sens, on demande des propriétés du rapport entre le langage et l'état d'affaires (modèle mathématique, collection de mondes possibles, etc.) qui s'appliquent à tous les énoncés (voire, toutes les énonciations) du langage ; elle n'est donc pas locale mais *globale*.

Etapas d'inférence dans un cadre local particulier, certains énoncés (et certains contrefactuels) sont « vrais », valides ou admis ; comprise de manière « inférentielles », une règle logique permet la « transition » de l'affirmation ou l'admission de certaines prémisses à l'affirmation ou l'admission de la conclusion : par conséquent, elle pourrait être comprise, dans l'optique locale, comme permettant une certaine *transition* de(s) cadre(s) local(ux) où la (les) prémisses est (sont) valide(s) à un cadre où la conclusion est valide. Vu sous cet angle, une règle logique correspondrait non pas à un aspect de la structure *dans* le cadre, mais à une opération *sur* les cadres, qui est dans un certain sens *permissible*.

La considération dans cette Annexe s'organise ainsi. Dans la première section, on présente les règles admises comme fondamentales dans la plupart des

logiques dites non monotones. Dans la mesure où les logiques des contrefactuels sont des logiques non-monotones, ces règles sont également admises par la plupart des théories de contrefactuels. Les trois sections suivantes portent respectivement sur les propriétés des cadres (locaux), les opérations sur les cadres, et les propriétés de la sémantique globale auxquelles peuvent référer les règles logiques. Dans chaque cas, la discussion sert à une double fin : d'une part, mettre en évidence les différentes interprétations des règles logiques, d'autre part, montrer le rapport avec les systèmes logiques proposés dans la tradition, tantôt en les reproduire dans le cadre proposé, tantôt en les comparer aux aspects du système proposé. Puisque c'est la forme plutôt que les détails des résultats qui importe ici, on omet les démonstrations détaillées dans cette Annexe, en présentant au plus des esquisses des démonstrations les plus complexes.

A.1 La logique non monotone aujourd'hui

Les règles de la logique non monotone

Le contrefactuel est reconnu depuis longtemps comme un exemple d'une conséquence dite non monotone : *non* monotone au sens où, alors que $A \Rightarrow B$, il n'est pas nécessairement le cas pour tout C que $A \& C \Rightarrow B$ (voir l'exemple dans 4.2.1). Par conséquent, la « logique » des contrefactuels est un exemple d'une logique dite non monotone. On traite ici la logique non-monotone, et donc, si l'on peut dire, la « conséquence non monotone », en général ; on parlera ainsi des conséquences plutôt que des contrefactuels, et on emploie le même symbole \Rightarrow pour elles. Il s'agit de considérer le système proposé par Kraus, Lehmann & Magidor, 1990¹ qui est souvent regardé comme le système « standard ». Ce système consiste en un langage propositionnel ordinaire \mathcal{L} – avec \rightarrow pour l'implication matérielle – et une opérateur métalinguistique \Rightarrow (pour la conséquence non monotone), qui satisfait aux règles suivantes :

¹Voir Kraus et al. (1990, §3.1).

$\overline{A \Rightarrow A}$	(R éfléxivité)
$\frac{A \leftrightarrow B \quad A \Rightarrow C}{B \Rightarrow C}$	(E quivalence L ogique sur la G auche)
$\frac{A \rightarrow B \quad C \Rightarrow A}{C \Rightarrow B}$	(A ffaiblissement sur la D roite)
$\frac{A \wedge B \Rightarrow C \quad A \Rightarrow B}{A \Rightarrow C}$	(C oupure, ou Transitivité Cumulative)
$\frac{A \Rightarrow B \quad A \Rightarrow C}{A \wedge B \Rightarrow C}$	(M onotone P rudente)

On ajoute souvent la règle suivante pour « renforcer » la logique.

$$\frac{A \Rightarrow C \quad B \Rightarrow C}{A \vee B \Rightarrow C} \quad (\mathbf{O}r)$$

La sémantique

On a proposé une « sémantique » pour une logique comportant ces six règles, sous forme d'un « modèle préférentiel » :

Définition A.1.1 (Modèle préférentiel²). Un *modèle préférentiel* pour un langage propositionnel \mathcal{L} (comportant \wedge, \vee, \neg), et une notion de consistance sur les énoncés de \mathcal{L} , est un triplet (S, l, \prec) , où

- S est un ensemble, dont on appelle les membres les états ;
- $l : S \rightarrow \wp(\mathcal{L})$ assigne à chaque état un ensemble d'énoncés du langage (les énoncés vrais dans cet état), de telle sorte que la valuation $v_s : \mathcal{L} \rightarrow \{\top, \perp\}$ de \mathcal{L} définie comme $v_s(A) = V$ si $A \in l(s)$; $v_s(A) = F$ si $A \notin l(s)$ est consistante pour tout $s \in S$.
- \prec est un ordre sur S qui est

irréflexif $s \not\prec s$

transitif si $s \prec t$ et $t \prec u$, alors $s \prec u$

²Kraus et al. (1990, §5.2)

lisse (ou finement bouchée) $\forall A \in \mathcal{L}, \hat{A} = \{s \in S \mid A \in l(s)\}$ a la propriété suivante (elle est *lisse*) : $\forall t \in \hat{A}$, soit $\exists s$ minimal $_{\prec}$ dans \hat{A} , $s \prec t$, soit t est minimal $_{\prec}$ dans \hat{A} .

où un état est minimal $_{\prec}$ s'il est minimal selon l'ordre \prec . On dirait dans la suite d'un état qui est minimal $_{\prec}$ dans \hat{A} qu'il est *minimal $_{\prec}$ -A*.

Les états de S sont intuitivement des « mondes possibles » – des « mondes » où la valeur de vérité de chaque énoncé de \mathcal{L} est déterminée ; ainsi, on écrit parfois $s \models A$ lorsque $A \in l(s)$. Pourtant, plusieurs états pourront valider exactement les mêmes énoncés ; autrement dit, un « même » monde pourrait prendre deux positions différents selon l'ordre \prec .

Chaque modèle préférentiel définit une relation de conséquence :

Définition A.1.2. Pour un modèle préférentiel $W = (S, l, \prec)$, la relation de conséquence \Rightarrow_W définie par le modèle W est telle que $A \Rightarrow_W B$ ssi pour tout s minimal dans \hat{A} , $s \models B$.

L'intérêt de ces définitions tient au résultat suivant :

Théorème A.1.1 (Théorème de Représentation³). *Une relation de conséquence \Rightarrow_W satisfait les six règles ci-dessus si et seulement si elle est la relation définie par un modèle préférentiel.*

C'est à dire que le modèle préférentiel donne une sémantique complète pour le système compris de règles **(Réfl)-(Or)** : pour chaque relation de conséquence qui satisfait à ces règles, il y a un modèle préférentiel qui la définit⁴.

A.2 Règles logiques comme propriétés des cadres

Selon la première interprétation des règles logiques dans le cadre proposé, elles renvoient à la structure des cadres locaux où les contrefactuels sont évalués (à des moments et dans des circonstances particuliers). Il s'agit, dans cette section, de ce que l'on a appelé la « sémantique locale ».

Les deux aspects de la théorie « standard » de la logique non monotone présentée dans §A.1 – les règles d'inférence et le modèle préférentiel – apparaissent comme deux manières de caractériser l'évaluation des conséquences dans un cadre local. Les règles s'appliquent directement à la conséquence définie par

³Kraus et al. (1990, Theorem 3).

⁴Makinson (2005, §2.1).

le cadre local, c'est-à-dire à l'ensemble de conséquences $A \Rightarrow B$ qu'il valide. En revanche, le modèle préférentiel, dans la mesure où il « caractérise » les règles qui sont satisfaites par la conséquence définie par le cadre local (selon le Théorème de Représentation), devrait entretenir une relation étroite avec la structure du cadre local qui définit cette conséquence. La logique non monotone, comprise comme sémantique locale, propose donc une structure générale pour le cadre local (le modèle préférentiel) et une collection de règles logiques qui « correspondent » exactement à, ou qui « caractérisent », cette structure. En particulier, elles *ne sont pas entendues* comme des « pas » ou des « étapes » dans des preuves où l'on déduit des vérités relatives aux conséquences – c'est l'interprétation des règles d'inférences comme des transitions entre cadres, discutée dans la section suivante, qui répond à cette notion de règle – mais plutôt à des propriétés structurelles et statiques des cadres locaux individuels.

En revanche, une large partie de l'argumentation des Chapitres 4 et 5 vise à proposer une structure pour le cadre local. Cette structure consiste en un cadre au sens de §5.1.2, c'est-à-dire une algèbre interprétée et un ensemble co-indépendant. Il s'avère que cette structure, et la conséquence qu'elle définit, diffère de la structure et de la logique standard. Il importe de le souligner : l'argumentation visait seulement à proposer une *sémantique locale*, et s'oppose à la logique standard seulement entendue comme *sémantique locale*. Aucune position n'a été prise sur la question des règles logiques qui s'appliquent à la transition entre cadres (aux étapes d'un raisonnement où figurent des contrefactuels), ou à une sémantique globale ; comme on verra dans les sections suivantes, on peut reproduire la logique standard lorsqu'il est comprise dans ces acceptions.

Il y a trois manières dont le système « standard » présenté dans §A.1 diffère de la structure proposée dans les chapitres précédents.

Premièrement, le système présuppose un langage fixe, alors que le cadre défini est local, en ce sens que, s'il y a un langage global fixe, seulement certain de ces énoncés figure dans chaque cadre. Par conséquent, l'analyse (**ALP**) ne valide pas certaines instances des règles de la logique non monotone standard en raison de l'absence des énoncés : la règle de réflexivité ((**Réf**)), par exemple, n'est pas validée pour un énoncé A et un cadre (\mathbf{B}, E) si A ne figure pas dans \mathbf{B} , car ce cadre ne permet même pas d'évaluer cette conséquence.

Deuxièmement, le système standard évalue chaque conséquence qui peut être formulé dans le langage (fixe), alors que, même pour le langage local en jeu, les contraintes de l'analyse (**ALP**) (§5.1.3) impliquent que seulement certaines conséquences qui peuvent être formulées dans ce langage local peuvent

être évalués. Pour reprendre l'exemple de la règle de réflexivité, la conséquence $A \Rightarrow A$ ne sera pas évaluée si A figure dans \mathbf{B} , mais A n'appartient pas à l'ensemble co-indépendant E .

Finalement, au-delà de ces deux différences, la structure du cadre proposé, c'est-à-dire la structure consistant en une algèbre interprétée et un ensemble co-indépendant, est plus riche que la structure du modèle préférentiel (Définition A.1.1), de telle sorte que la conséquence définie ou impliquée par le premier a quelques propriétés supplémentaires relatives à celle, standard, qui est définie par le dernier.

La première différence relève du fait que, dans la logique non monotone standard il n'y a pas de notion de cadre locale et donc de distinction entre la logique en tant qu'elle s'applique dans ou hors de ces cadres. Néanmoins, cette différence est surmontée par hypothèse, dans la mesure où l'on considère le système de logique non monotone standard en tant qu'éventuelle proposition de système pour décrire la structure interne des cadres locaux, ou, pour reprendre le terme introduit ci-dessus, comme « sémantique locale ». Sous cette interprétation, on pourrait supposer un langage local fixe, commun au système non monotone standard et au système proposé ici. Il s'agit effectivement de travailler dans une algèbre interprétée fixe, ou de manière équivalente, un langage (local) fixe avec une interprétation fixe dans un ensemble de (petits) mondes possibles (§§4.2.1 et 5.1.1). La comparaison est donc entre l'analyse offerte par le genre d'ordre proposé par les modèles préférentiels (Définition A.1.1) et celle offerte par l'ensemble co-indépendant (Définition 5.1.7).

La deuxième différence appartient aux particularités de l'analyse (**ALP**), et notamment aux contraintes d'indépendance de l'antécédent par rapport aux prémisses (compte tenu du conséquent) et de pertinence de l'antécédent par rapport au conséquent. Comme l'a été remarqué dans §5.2.1, cette analyse pourrait être modifiée pour affaiblir ces contraintes. Dans la mesure où le système standard, dont la ressemblance à la théorie des contrefactuels proposée par Lewis et Stalnaker est évidente et bien connue, a moins de scrupules au sujet de pertinence et d'indépendance, comme on l'a d'ailleurs constaté relativement à la théorie de Lewis-Stalnaker dans §4.2.3, pour bien apercevoir la différence de structure entre le modèle préférentiel et l'ensemble co-indépendant, il conviendrait de comparer le système standard avec une version affaiblie de l'analyse (**ALP**) qui le dépouille de ces contraintes. Dans §A.2.1 on définira une notion affaiblie de satisfaction sur les algèbres interprétées équipées d'ensembles co-indépendants, qui permet d'évaluer tous les conséquences qui peuvent être

formulée dans le langage local de cette algèbre (Définition A.2.1) et donc de mieux comparer les deux analyses.

Or, même avec ces deux réserves, la structure du cadre proposé dans §5.1 diffère de celle des modèles préférentiels proposée dans la tradition, de telle sorte que les conséquences qu'ils définissent ont des propriétés différentes. Il y a deux manières de mettre en évidence ces différences. Premièrement, on peut comparer les deux structures ou modèles proposés, à savoir les ensembles co-indépendants et les cadres. Pour ce faire, on « traduira » le formalisme d'ensembles co-indépendants dans le formalisme d'ordres employés dans la définition de modèles préférentiels. Plus précisément, dans §A.2.2, on définit une notion de modèle-**(ALP)**, qui consiste en un ordre comme celui impliqué dans le modèle préférentiel mais avec quelques propriétés supplémentaires, et qui correspond à un ensemble co-indépendant, au sens où il définit un ensemble co-indépendant qui valide les mêmes conséquences. Deuxièmement, de même que, dans le système traditionnel, il y a une collection de règles qui caractérisent la conséquence non monotone définie par les modèles préférentiels (d'après le Théorème de Représentation ; Théorème A.1.1), on pourrait demander les règles qui caractérisent la conséquence définie par les algèbres interprétées et les ensembles co-indépendants. Dans §A.2.3, on présente de telles règles, et formule des théorèmes de représentation à leur égard, pour le cas de l'analyse originale (**(ALP)**) et l'analyse affaiblie.

A.2.1 Préliminaires : une analyse relâchée

L'analyse **(ALP)** (§5.1.3) permet, pour la donnée d'une algèbre interprétée **B** et un ensemble interprété E de cette algèbre, d'évaluer toutes les conséquences dont l'antécédent A est membre de E et le conséquent C est membre de $B_E \setminus B_{E \setminus \{A\}}$ ⁵. Pour étendre l'évaluation aux conséquences dont le conséquent est membre de B_E , il suffit de lever la condition que $C \in B_E \setminus B_{E \setminus \{A\}}$ (condition 3. de l'analyse, ou l'exigence que l'antécédent soit pertinent au conséquent). Pour étendre l'évaluation aux conséquences où l'antécédent est membre de B_E , il suffit de constater d'abord qu'elle s'étend facilement aux cas où l'antécédent est une conjonction des éléments de E , et ensuite de rappeler le résultat bien connu que chaque élément de B_E pourrait être écrit de manière unique comme la disjonction des conjonctions des éléments de E (en dite forme normale disjonctive-conjonctive). Cela donne la notion de satisfaction suivante de conséquences relative à un cadre (\mathbf{B}, E) .

⁵ B_E est l'algèbre libre engendré par E : voir Définition 5.1.7.

Définition A.2.1 (Satisfaction_r sur un cadre). Un cadre (\mathbf{B}, E) avec une valuation v des éléments de E satisfait_r une conséquence $P_1 \wedge \dots \wedge P_l \Rightarrow B$ où $B \in B_E$ et, pour tout $i \in \{1, \dots, l\}$, $P_i \in E$ ssi $P_1 \wedge \dots \wedge P_l \wedge Q_1 \wedge \dots \wedge Q_k \leq B$, où pour tout $j \in \{1, \dots, k\}$, $Q_j \in E \setminus \{P_1, \dots, P_l\}$, et $v(Q_j) = V$.⁶

Un cadre (\mathbf{B}, E) avec une valuation v des éléments de E satisfait_r une conséquence $A \Rightarrow B$ ssi $A, B \in B_E$ et $A = \bigvee_j \bigwedge_{P_i^j \in E} P_i^j$ avec $\bigwedge P_i^j \Rightarrow B$ pour tout j .

A.2.2 L'analyse (ALP) en termes d'ordres

Il s'agit dans cette section de définir un type d'ordre, à la façon des ordres impliqués dans les modèles préférentiels, tels que les ensembles d'états ou de mondes (ou, de manière équivalente, les algèbres interprétées) équipées des ordres de ce type donnent la même évaluation des conséquences qu'une algèbre interprétée équipée d'un ensemble co-indépendant. On définit d'abord la notion de modèle-(ALP) comportant ce genre d'ordre, et ensuite deux notions de satisfaction d'une conséquence sur ce modèle, dont la première correspond à la satisfaction ordinaire selon l'analyse (ALP), et la dernière à la satisfaction_r définie ci-dessus.

Le modèle-(ALP)

Le type de « modèle » qui correspond au cadre (Définition 5.1.8) est le suivant.

Définition A.2.2 (Modèle-(ALP)). Un *modèle-(ALP)* W pour un langage propositionnel L^7 est la donnée de (S, l, \sim, \prec) , où

- S est un ensemble d'états;
- $l : S \rightarrow \wp(L)$ assigne à chaque état une valuation consistante;
- \sim est une relation d'équivalence sur les états de S avec $\exists x_W \in N |\{S_\sim\}| = 2^{x_W}$ (le nombre de classes d'équivalence de S sous \sim est une puissance de 2). On appelle x_W la « largeur » de W .
- \prec est un ordre sur S_\sim tel que

irréflexif $[s]_\sim \not\prec [s]_\sim$;

structuré de treillis la clôture réflexive de \prec est un treillis distributif sur S_\sim . L'élément minimal du treillis est appelé $[v_W]_\sim$. On définit le

⁶Rappeler la Remarque 5.1.2 portant sur le fait que la notion d'ensemble co-indépendant est employée à la négation près.

⁷On emploie L plutôt que \mathcal{L} pour indiquer que, puisque l'on travaille dans des localités relatives à des instants donnés, le langage est « local ».

rang d'un élément $r([s]_{\sim})$ comme le n maximal tel qu'il y a une suite $[v_W]_{\sim} \prec [s_1]_{\sim} \prec \cdots \prec [s_n]_{\sim} = [s]_{\sim}$;

respectant de la largeur $|\{[u]_{\sim} \mid r([u]_{\sim}) = n\}| = \binom{x_W}{n}$.

Pour un élément $[s]_{\sim}$ de la partition, on écrit $[s]_{\sim} \models A$ si $\forall t \in S$, si $t \in [s]_{\sim}$ alors $A \in l(t)$. On dit qu'un énoncé $A \in L$ *respecte* la partition de W si $\forall [s]_{\sim} \in S_{\sim} [s]_{\sim} \models A$ ou $[s]_{\sim} \models \neg A$.

Observation A.2.1. L'ordre \prec est irréflexif, transitif et lisse (ou finiment bouché), de telle sorte que, selon la Définition A.1.1, un modèle-**(ALP)** est un modèle préférentiel sur les éléments de la partition par \sim (les classes d'équivalences de \sim)⁸

La relation d'équivalence répartit les « (petits) mondes » ou « états » du modèle ; la relation \prec opère au niveau de cette partition, plutôt que celle des « mondes » eux-mêmes. En termes des algèbres interprétées et des ensembles co-indépendants, les éléments de S_{\sim} seront les éléments de B_E . L'ordre \prec dicte qu'un élément de la partition (élément de B_E) est plus éloigné qu'un autre si, par rapport à la valuation des éléments de E (dans la « réalité ») il faut « modifier la valeur » de plus d'éléments de E pour que atteindre le premier élément que pour atteindre le dernier⁹. D'où la définition suivante de satisfaction d'une conséquence sur un modèle-**(ALP)**.

Définition A.2.3 (Satisfaction sur un modèle-**(ALP)**). Pour A et B dans le langage local L d'un modèle-**(ALP)** W , W satisfait conséquence $A \Rightarrow B$ si

1. A et B respectent la partition de W ;
2. $\exists! [s]_{\sim}$ tel que, si $[v_W]_{\sim} \models A$, $[s]_{\sim} = [v_W]_{\sim}$, et si $[v_W]_{\sim} \models \neg A$ $r([s]_{\sim}) = 1$ et $[s]_{\sim} \models A$;
3. il n'existe pas de bijection involutive $\phi : (S_{\sim}, \prec) \rightarrow (S_{\sim}, \prec)$ avec
 - $[s]_{\sim} \models A$ ssi $\phi([s]_{\sim}) \models \neg A$;
 - $r([s]_{\sim}) = r(\phi([s]_{\sim})) \pm 1$;
 - $[s]_{\sim} \models B$ ssi $\phi([s]_{\sim}) \models B$.

Remarque A.2.1 (Correspondance : l'intuition). Intuitivement, ces conditions correspondent respectivement aux conditions suivantes de l'Analyse **(ALP)** (§5.1.3).

⁸Cette observation vaut indépendamment de la taille de S . Or, dans cette section et dans la prochaine, où il s'agit des structures logiques locales, la plupart des algèbres interprétées (et donc des ensembles S) sont effectivement finies, si bien que certaines contraintes sur l'ordre, dont le fait d'être lisse, sont forcément satisfaites.

⁹Cet ordre a été employé dans la discussion de l'exemple de Lewis, §5.3.1.

1. correspond à l'exigence que A et B appartiennent à B_E (conséquence de la Condition 1., première partie de la Condition 3.) ;
2. correspond à l'exigence que $A \in E$ est que l'on a affaire aux P_i vrais (Condition 2.) ;
3. correspond à la pertinence de A par rapport à B (Condition 3.).

L'autre analyse, avec la notion de satisfaction affaiblie (satisfaction_r , §A.2.1), correspond à la notion de satisfaction suivante.

Définition A.2.4 (Satisfaction_r sur un modèle-(ALP)). Un modèle-(ALP) W satisfait $_r$ une conséquence $A \Rightarrow B$ ssi A et B respectent la partition de W et pour tout $[s]_{\sim} \text{minimal}_{\prec} A^{10}$, $[s]_{\sim} \Vdash B$.

Observation A.2.2. La notion de satisfaction_r coïncide avec la satisfaction sur les modèles préférentiels définie dans Définition A.1.2 (au niveau de la partition selon \sim , ou de manière équivalent, des éléments de B_E ; voir l'Observation A.2.1).

Équivalence de cadres et de modèles-(ALP)

Les traductions suivantes précisent le rapport entre les cadres et les modèles-(ALP).

Définition A.2.5 (Traduction de cadres aux modèles-(ALP)). Soit (\mathbf{B}, E) un cadre local ($\mathbf{B} = (B_I, B, q)$) et $v : E \rightarrow \{V, F\}$ une valuation de tous les éléments de E . On définit un modèle-(ALP) engendré par (\mathbf{B}, E) , $W_{(\mathbf{B}, E)} = (S, l, \sim, \prec)_{(\mathbf{B}, E)}$, pour le langage propositionnel $L_{(\mathbf{B}, E)} = B_I$, de manière suivante :

- S est l'ensemble d'atomes de B ;
- $l(s) = \{A \in \mathbf{B} \mid s \leq q(A)\}$;
- \sim est la relation d'équivalence sur les états de S définie par $s \sim t$ ssi $\{A \in E \mid A \in l(s)\} = \{A \in E \mid A \in l(t)\}$;
- \prec est l'ordre sur S_{\sim} définie par $[s]_{\sim} \prec [t]_{\sim}$ ssi $\{A \in E \mid [s]_{\sim} \Vdash A \ \& \ v(A) = F \text{ ou } [s]_{\sim} \Vdash \neg A \ \& \ v(A) = V\} \subsetneq \{A \in E \mid [t]_{\sim} \Vdash A \ \& \ v(A) = F \text{ ou } [t]_{\sim} \Vdash \neg A \ \& \ v(A) = V\}^{11}$.

Il n'est pas difficile de montrer que cette définition est bonne, c'est-à-dire que $W_{(\mathbf{B}, E)}$ satisfait les conditions de la Définition A.2.2.

¹⁰ $\text{minimal}_{\prec} A$ est définie dans la Définition A.1.1, à ceci près qu'ici \hat{A} est un ensemble d'éléments de la partition plutôt que d'états de S .

¹¹Rappeler (Remarque 5.1.2) que l'ensemble co-indépendant est effectivement employé à la négation de ses éléments près.

Définition A.2.6 (Traduction de modèles-(ALP) aux cadres). Soit $W = (S, l, \sim, \prec)$ un modèle-(ALP) pour un langage propositionnel L . On définit un cadre $(\mathbf{B}, E)_W$ et une valuation $v_W : E \rightarrow \{V, F\}$ de manière suivante :

- I est l'ensemble d'énoncés de L ;
- \mathcal{F} est l'ensemble de formules ϕ figurant les éléments de I telles que, pour tout $s \in S$, $\phi \in l(s)$;
- $\mathbf{B}_W = (B_I, B_I/\mathcal{F}, q)^{12}$;
- $E = \{A \in \mathbf{B}_W \mid [v_W]_{\sim} \Vdash A \ \& \ \exists [s]_{\sim} (r([s]_{\sim}) = 1 \ \& \ [s]_{\sim} \Vdash \neg A)\}$;
- pour tout $P \in E$, $v_W : P \mapsto V^{13}$.

Il n'est pas difficile de montrer que cette définition est bonne, c'est-à-dire que $(\mathbf{B}, E)_W$ satisfait les conditions de la Définition 5.1.8.

Avec ces traductions, il y a une équivalence entre l'évaluation d'une conséquence dans un cadre et la satisfaction dans un modèle-(ALP).

Proposition A.2.1. *Si $A \Rightarrow B$ est satisfaite dans un cadre (\mathbf{B}, E) , alors elle est satisfaite dans $W_{(\mathbf{B}, E)}$. À l'inverse, si elle est satisfaite dans un modèle-(ALP) W , alors elle est satisfaite dans $(\mathbf{B}, E)_W$.*

En outre, la satisfaction_r sur les cadres est équivalente à la satisfaction_r sur les modèles-(ALP).

Proposition A.2.2. *Si $A \Rightarrow B$ est satisfaite_r dans un cadre (\mathbf{B}, E) , alors elle est satisfaite_r dans $W_{(\mathbf{B}, E)}$. À l'inverse, si elle est satisfaite_r dans un modèle-(ALP) W , alors elle est satisfaite_r dans $(\mathbf{B}, E)_W$.*

Observation A.2.3. Il s'ensuit immédiatement de cette Proposition, du Théorème de Représentation pour les modèles préférentiels (Théorème A.1.1), et des Observations A.2.1 and A.2.2, que l'analyse en termes de cadres employant la satisfaction_r satisfait aux règles (Réfl)-(Or) (relatif aux B_E).

A.2.3 Règles pour les analyses proposées

Dans la perspective de la sémantique locale, le Théorème de Représentation exprime l'équivalence entre un certain nombre de contraintes sur, ou propriétés de, une relation de conséquence – contraintes sur l'ensemble de conséquences qui sont valides (dans un cadre local donné) – et une structure logique (celle du

¹²Rappeler (§5.1.1) qu'une définition alternative d'une algèbre interprétée est comme le quotient de l'algèbre d'interprétation par un ensemble de formules exprimées en termes de ses éléments.

¹³On s'appuie sur le fait que, dans la clause précédent, on a choisit comme représentant des éléments de E (qui sont strictement des classes d'équivalences sous la négation), le représentant qui est évalué comme vrai.

cadre) qui permet d'évaluer la validité de ces conséquences. Comme remarqué ci-dessus (Observations A.2.1 et A.2.2), la structure du cadre local ou du modèle-**(ALP)** est plus forte que celle du modèle préférentiel, si bien que, pour retrouver à nouveau une équivalence entre les contraintes sur les conséquences validées et la structure logique où ils sont évalués, il faudrait ajouter de nouvelles contraintes sur l'ensemble de conséquences – de nouvelles « règles », si l'on préfère. C'est le cas aussi bien pour la notion de satisfaction employée dans l'analyse **(ALP)** que pour la notion relâchée de satisfaction_r (Définition A.2.1).

Règles pour l'Analyse (ALP)

Soit L un langage propositionnel L , avec des opérateurs $\wedge, \vee, \neg, \rightarrow$, et donc \leftrightarrow (bi-implication). Dans le cas (local) présent, on suppose que le langage, et sa notion d'équivalence logique, est telle que l'algèbre de Boole formée, B_L est finie¹⁴.

Pour une relation conséquence définie par (la notion de satisfaction impliquée dans) l'Analyse **(ALP)**, certains énoncés A sont des antécédents des conséquences valides $A \Rightarrow B$, alors que certains peuvent ne pas l'être. Soit $Eval$ l'ensemble des A qui sont des antécédents d'au moins une conséquence valide. Une première propriété, ou « règle » qui est satisfaite par la relation de conséquence impliquée par **(ALP)**, est la suivante.

$$\begin{array}{ll} \text{si } A \in Eval, \text{ alors } \neg A \in Eval & (\text{Clôture sous} \\ \text{si } A \in Eval, A \leftrightarrow B \text{ alors } B \in Eval & \text{Négation et Équivalence}) \end{array}$$

Par suite de cette règles, on pourrait travailler avec l'ensemble $E = Eval / \sim'$, où $A \sim' B$ ssi $A \leftrightarrow B$ ou $A \leftrightarrow \neg B$. Pour un énoncé $A \in L$, on utilise $A \in E$ pour signifier que A est un représentant d'une classe dans E .

Il s'agira des « règles » ou contraintes suivantes sur une relation de conséquence.

¹⁴On peut enlever cette supposition en ajoutant une règle exigeant que $Eval$ est finie.

$$\forall A_1, \dots, A_n \in Eval \ A_i \approx' A_j \text{ pour } i \neq j, \nexists CL : \quad (\mathbf{Co-Indépendance})$$

$$L^n \rightarrow L, tq. CL(A_1, \dots, A_n) = \perp$$

(Tiers Exclu Restreint)

$$\forall A \in Eval \ \exists B \in L, (B \leftrightarrow \bigwedge_{\substack{P_i \in E \\ 1 \leq i \leq |E|}} P_i) \text{ et } (A \Rightarrow B)$$

Où CL prend une suite des énoncés à une certaine combinaison logique de ces énoncés.

Remarque A.2.2. La troisième règle s'appelle le Tiers Exclu Restreint, car elle implique la propriété suivante, pour tout A qui figure comme antécédent d'une conséquence valide :

$$\forall B \in L \ tq. B \leftrightarrow \bigvee \bigwedge_{\substack{P_i \in E \\ 1 \leq i \leq |E|}} P_i, \quad \text{soit } \exists C, A \Rightarrow C \text{ et } C \rightarrow B$$

$$\text{soit } \exists C', A \Rightarrow C' \text{ et } C' \rightarrow \neg B$$

La différence entre cette propriété et le tiers exclu pour la conséquence (voir par exemple Stalnaker (1980)) tient au fait que l'Analyse (**ALP**) ne s'exprime pas sur la validité des conséquences dont le conséquent n'est pas combinaison logique des antécédents possibles ou où il est indépendant de l'antécédent (Condition 3. de l'Analyse (**ALP**)).

Il convient de rappeler que, dans la perspective de la sémantique locale, dont on s'occupe dans cette section, les règles ne sont pas conçus comme des « pas » permis dans une démonstration, mais plutôt des propriétés ou des contraintes sur la structure du cadre local, exprimées uniquement en termes de propriétés des conséquences qui y sont validées. Les règles qui viennent d'être proposées sont toujours *syntaxiques*, dans la mesure où elles s'expriment uniquement en termes des énoncés du langage et des rapports entre eux ; en ce sens, elles sont bien des « règles logiques » au même titre que les règles proposées traditionnellement (§A.1)¹⁵.

Soient (**Réfl'**) la règle (**Réfl**) s'appliquant seulement aux cas où $A \in Eval$, et (**AD'**) la règle (**AD**) s'appliquant seulement aux cas où $B \leftrightarrow \bigvee \bigwedge_{\substack{P_i \in E \\ 1 \leq i \leq |E|}} P_i$.

¹⁵La différence apparente entre ces règles et celles de § A.1 tient à la possibilité de les exprimer *localement* – relatives à des énoncés individuels et les rapports entre eux – plutôt que *globalement* – relatives à des ensembles des énoncés (toutes les conséquences).

Théorème A.2.1 (Théorème de Représentation pour Analyse (ALP)). *Une relation de conséquence \Rightarrow_W dans un langage L satisfait les règles **(Réfl')**, **(AD')**, **(ELG)**, **(CNE)**, **(Co-Ind)** et **(TER)** si et seulement si elle est la relation définie par un cadre et une valuation des éléments de son ensemble co-indépendant, avec la notion de satisfaction d'Analyse **(ALP)**.¹⁶*

Règles pour l'Analyse (ALP)_r

En supposant toujours un langage propositionnel L , avec B_L finie, soit Ant l'ensemble des A qui sont des antécédents d'au moins une conséquence valide, selon la notion de satisfaction_r et $Cons$ l'ensemble de C qui sont des conséquents d'au moins une conséquence valide. Ant est ordonné par \rightarrow ; soit $Min \subset Ant$ l'ensemble d'énoncés minimaux sous cet ordre. On a tout de suite la règle suivante

Ant est clos sous les opérations logiques (\neg, \wedge) (Clôture Logique)

de sorte que l'on peut travailler avec $At = Min / \leftrightarrow$ (comme avant, $A \in At$ signifie que A représente un élément de At).

Soient les « règles » ou contraintes structurelles suivantes.

$Ant = Cons$ (Saturation)

$|At| = 2^x$ pour un $x \in N$ (Algèbre Libre)

$|\{A \in L \mid \exists B \in At, A \Rightarrow B \ \& \ \neg \exists C (A \rightarrow C \ \& \ C \Rightarrow B) \ \& \ |\{C \in At \mid C \rightarrow A\}| = 2^{x-n}\}| / \leftrightarrow = \binom{x}{n}$ (Respectant Largeur)

On pourrait répéter ici les remarques précédentes relatives au caractère *syntactiques* de ces règles. Le résultat cherché est le suivant (où **(Réfl')**, **(AD')** sont les restrictions des règles appropriées à Ant).

Théorème A.2.2 (Théorème de Représentation pour Analyse (2)_r). *Une relation de conséquence \Rightarrow_W dans un langage L satisfait les règles **(Réfl')**, **(AD')**, **(ELG)**, **(Coup)**, **(MP)**, **(Or)**, **(S)**, **(CL)**, **(AL)** et **(RL)** si et seulement si elle est la relation définie par un modèle-**(ALP)** avec la notion de satisfaction_r.¹⁷*

¹⁶Esquisse de la démonstration : on prend $B_I = B_L$; $B = B_L / \leftrightarrow$; l'ensemble co-indépendant E contenant un représentant de chaque élément de E ; et, si A, B représentent deux éléments différents de E , mais $\exists C, C \leftrightarrow \bigwedge_{\substack{P_i \in E \\ 1 \leq i \leq |E|}} P_i, A \Rightarrow C$ et $B \Rightarrow C$ alors, pour $D \in Eval, v(D) = T$ ssi $C \rightarrow D$.

¹⁷Esquisse de la démonstration : on prend $B_I = B_L$; $B = B_L / \leftrightarrow$; $E = \{A \in L \mid \exists B \in$

Remarque A.2.3. Ce théorème confirme l'Observation A.2.3, selon laquelle la relation de conséquence définie par les cadres est plus forte que celle définie par le modèle préférentiel : elle satisfait à des règles supplémentaires.

A.3 Règles logiques comme règles de transition entre cadres

Parfois on entend des règles logiques comme des recettes permettant (ou interdisant) des « *pas* » ou des *transitions* d'un certain nombre d'énoncés à d'autres. Dans la perspective de la logique locale, il s'agit non plus des propriétés internes aux cadres locaux, mais plutôt des propriétés ou contraintes sur les transitions entre, ou des opérations sur des cadres locaux. La théorie de contrefactuels proposée dans la seconde partie de cette thèse n'a pas pris position sur la question de telles propriétés, ou de telles contraintes ; néanmoins, comme on verra prochainement, le cadre théorique qu'elle propose permet de définir de manière très naturelle des opérations sur les cadres qui satisfont les règles traditionnelles de la logique non monotone (§A.1). Ceci devrait être compris comme un atout technique du cadre théorique et conceptuel proposé, à savoir une certaine diversité, plutôt qu'une sanction de ces règles. D'autant plus que ces opérations sont des cas spéciaux des relations entre cadres locaux, et les règles logiques sont satisfaites de manière naturelle seulement dans des cas assez particuliers.

D'ailleurs, on a déjà mentionné des raisons pour douter des règles logiques strictes proposées pour les contrefactuels, et *a fortiori* pour la logique non monotone en général. Dans §4.3.1, on a donné d'apparents contre-exemples à certaines règles de §A.1, en l'occurrence :

(Coup) « S'il pleuvait, j'apporterais une parapluie » pourrait être admissible, ainsi que « s'il pleuvait et j'apportais une parapluie, je ne me mouillerais pas » ; pourtant « s'il pleuvait, je ne me mouillerais pas » ne semble pas s'ensuivre.

(MP) « S'il pleuvait, je me mouillerais » et « s'il pleuvait, j'apporterais une parapluie » sont apparemment admissibles, alors que « s'il pleuvait et j'apportais un parapluie, je me mouillerais » ne semble pas s'ensuivre.

Il s'avérera que l'analyse des « étapes d'inférence » comme des opérations sur, ou des transitions entre, des cadres locaux permettra non seulement de re-

$At, A \Rightarrow B \ \& \ \neg \exists C (A \rightarrow C \ \& \ C \Rightarrow B) \ \& \ |\{C \in At \mid C \rightarrow A\}| = 2^{x-1} \} / \leftrightarrow ;$ et $v(A) = T$ ssi $V \rightarrow A$, où $V = \bigwedge P_i \in EP_i$.

produire les règles logiques, telles **(Coup)** et **(MP)** mais en plus d'*expliquer* ces contre-exemples (§A.3.4). Cette conception permet *non seulement* de comprendre les cas où les règles s'appliquent, voire l'intuition qui les soutient, *mais également* les cas où elles ne s'appliquent pas de manière aussi évidente.

La discussion se déroule de manière suivante. Premièrement, l'idée d'interpréter, comprendre ou modéliser un sequent par un cadre ou structure locale sera présentée et illustrée sur l'exemple de la règle **(Réfl)**. Dans la mesure où les autres règles de §A.1 comportent deux prémisses et une conclusion, elles exigent une opération de « fusion » des cadres, selon le rapprochement entre sequents et cadres. On définit d'abord une notion de « fusion » des *algèbres interprétées* en termes des opérations de produit libre et de quotient. Cette « fusion » permettra déjà d'interpréter les règles **(ELG)**-**(AD)**. Ensuite, on étend la définition de la notion de « fusion » aux structures capables à évaluer les conséquences, en sorte de pouvoir analyser les règles **(Coup)**-**(Or)** et les contre-exemples mentionnés ci-dessus. Jusqu'à ce point on aura employé une structure simple pour cadre local, qui est plus faible que le cadre défini dans §5.1.2 (ou, de manière équivalente, le modèle-**(ALP)**, Définition A.2.2); on s'attardera enfin sur les difficultés d'étendre l'analyse aux cadres de §5.1.2, et leurs conséquences. Tout d'abord, il convient de définir la notion de cadre_f qui s'avérera techniquement utile dans cette section.

A.3.1 Préliminaire : le cadre_f

Dans cette section, on travaillera pour l'essentiel avec une théorie simple de la conséquence, qui s'appuie sur une algèbre interprétée équipée d'un ordre irréflexif, transitif et lisse¹⁸ \prec sur certains de ces éléments (les classes d'une partition). Plus précisément, il s'agit d'employer une structure que l'on appelle cadre_f (cadre faible), défini comme suit.

Définition A.3.1 (Cadre_f). Un cadre_f , (\mathbf{B}, \prec) consiste en une algèbre interprétée \mathbf{B} , et un ordre irréflexif, transitif et lisse¹⁹ \prec sur un ensemble d'éléments S de B , où

$$\forall x, y \in S, \quad x \wedge y = \perp$$

$$\bigvee_{x_j \in S} x_j = \top$$

¹⁸Puisque, en pratique, tous les langages locaux, et par conséquent tous les cadres et cadre_f , qui sont considérés ici sont finis, tout tel ordre sera automatiquement lisse; on continue de spécifier cette propriété afin de garder explicite son importance particulière. Voir également la note 8 ci-dessus.

¹⁹Voir la Définition A.1.1.

Remarque A.3.1 (Partition). Les deux conditions sur l'ensemble S affirment effectivement que les éléments de S forment une partition des atomes de \mathbf{B} , à la façon du S_{\sim} dans la Définition A.2.2 (voir notamment les remarques qui suivent cette définition). On dit d'ailleurs d'un ensemble d'éléments qui satisfait ces propriétés qu'il *constitue une partition*. Reprenant la terminologie introduite dans la suite de la Définition A.2.2, on dit qu'un élément z de \mathbf{B} *respecte* la partition constituée par S_i s'il existe $x_1, \dots, x_n \in S_i$ avec $z = x_1 \vee \dots \vee x_n$. Et, s'appuyant sur la convention terminologique introduite dans §5.1.2, on appelle l'ensemble²⁰ d'énoncés de \mathbf{B} qui sont équivalents à des combinaisons logiques des éléments de S , B_S ²¹; autrement dit, c'est l'ensemble d'énoncés qui respectant la partition.

L'ordre sur un cadre_f ressemble donc à l'ordre sur un modèle-(ALP), au sens où dans les deux cas il s'agit d'un ordre sur une partition des éléments de l'algèbre interprétée. De la sorte, on pourrait définir une notion d'évaluation de conséquences dans un cadre_f (en l'occurrence, des conséquences formées des énoncés de B_S), qui ressemble à la notion de satisfaction_r sur les modèles-(ALP) (Définition A.2.4), qui est d'ailleurs proche à la notion de satisfaction ordinaire sur les modèles préférentiels (Définition A.1.2). Il s'agit encore de la vérité du conséquent dans les « états » minimaux satisfaisant l'antécédent. Formellement :

Définition A.3.2 (Satisfaction sur un cadre_f). Un $\text{cadre}_f (\mathbf{B}, \prec)$ satisfait une conséquence $A \Rightarrow B$ ssi A et B respectent la partition de W et pour tout $s \in S$ minimal _{\prec} - A ²², $s \leq B$.

Les deux remarques suivantes permettent de situer le cadre_f , et la relation de conséquence qu'elle définit, par rapport aux autres modèles et autres modèles considérés ci-dessus.

Remarque A.3.2 (Cadre_f et cadre). La sous-algèbre B_S est l'équivalent pour un cadre_f de la sous-algèbre B_E d'un cadre ordinaire. Or, le cadre_f (et par conséquent la relation de conséquence qu'elle définit) est plus faible que le cadre, dans la mesure où le cadre_f ne satisfait pas aux conditions lourdes sur le cadre qui assurent que E soit un ensemble co-indépendant, et donc que B_E soit une algèbre libre. Autrement dit, l'ordre sur le cadre_f ne satisfait à toutes les conditions sur l'ordre sur le modèle-(ALP) (restriction sur la taille de S , et les propriétés de \prec dans la Définition A.2.2; voir §A.2.2 sur l'équivalence entre les cadres et les modèles-(ALP)).

²⁰Plus précisément : la sous-algèbre interprétée.

²¹Formellement : c'est l'ensemble de A , $q(A) \in B_S$, où B_S est l'algèbre libre engendrée par S (Définition 5.1.2).

²²minimal _{\prec} - A est défini de la même manière que dans la Définition A.1.1.

Remarque A.3.3 (Cadre_f et modèle préférentiel). Pour permettre une comparaison intéressante entre les cadres_f et les modèles préférentiel, on fait abstraction de la question des partitions, et on emploie le terme « états » pour désigner aussi bien les éléments de la partition dans un cadre_f et les éléments de l'ensemble de base d'un modèle préférentiel (les S dans les Définitions A.3.1 et A.1.1). La différence entre les deux tient au fait que, pour un langage (local) fixe, un modèle préférentiel peut comporter plusieurs états qui accordent les mêmes évaluations à *tous* les énoncés du langage, alors que, pour toute paire d'états d'un cadre_f, il existe des énoncés qu'ils évaluent différemment. Comme l'ont montré Lehmann et Magidor (1992), cette différence est significative : pour un langage fixe, il y a des conséquences définies par un modèle préférentiel qui ne sont pas définies par aucun cadre_f (ou autrement dit, aucun modèle dont tous les états se différencient selon les énoncés qu'ils valident). Pour un langage fixe, le cadre_f est donc une structure strictement plus forte qu'un modèle préférentiel ; en particulier, comme le modèle préférentiel il satisfait les règles **(Réfl)-(Or)** (§A.1, Théorème A.1.1).

Or, comme l'ont fait remarquer certains²³, la différence entre les deux disparaît dès que l'on ne se limite pas à un langage fixe. Un modèle préférentiel comportant deux états qui valident les mêmes énoncés (d'un langage L) correspond exactement à un cadre_f avec un langage (local, c'est-à-dire algèbre d'interprétation) L' comportant deux copies d'un énoncé $A \in L$, de telle sorte que les deux états se différencient selon la *copie* de A qui y est vraie²⁴. Alors que dans une perspective « globaliste », de tels changements du langage pourraient inquiéter, dans la perspective « localiste » recommandée ici, où la gestion des langages (locaux) est un thème crucial, l'idée glisser entre langages est tout à fait naturel. D'ailleurs, non seulement le fait de permettre plusieurs « copies » de ce qui *intuitivement* le même énoncé dans le langage local est cohérente avec la définition des algèbres interprétées (Définition 5.1.3)²⁵, il est peut-être exigé par des considérations indépendantes, comme on verra dans les §§A.3.3 et A.3.4 ci-dessous.

Alors que la discussion qui suit est formulée en termes des cadres_f, il sera évident que la plupart des considérations, et notamment toute l'argumentation de §§A.3.2 et A.3.3, s'applique également aux cadres. On aura l'occasion de revenir sur ce point dans §A.3.5, où l'on discute explicitement le cas des cadres.

²³Makinson (2005, §3.3) par exemple.

²⁴C'est-à-dire, il y a un isomorphisme entre les ensembles d'états (les algèbres de base) du modèle préférentiel et du cadre_f qui préserve les ordres sur eux.

²⁵Par exemple : puisque I est un ensemble quelconque (d'énoncés), rien n'empêche qu'il contient deux éléments A_1 et A_2 qui sont « intuitivement » le même énoncé.

Par ailleurs, il s'agira à des moments de considérer plusieurs cadres_f différents. En conséquence de la localité des algèbres interprétées (§5.1.1), il n'y a pas *a priori* d'identification des énoncés entre les algèbres interprétées, et donc des cadres_f, différentes²⁶. Pour modéliser l'identification des énoncés entre des différents cadres_f là où il en est besoin, on emploiera une relation de congruence \sim sur l'union disjoint des éléments des cadres qui relie seulement des éléments différents appartenant aux cadres différents²⁷. Donc, s'il s'agit de deux cadres_f \mathbf{B}_1 et \mathbf{B}_2 , on prendra une relation de congruence \sim qui relie seulement des éléments de \mathbf{B}_1 avec des éléments de \mathbf{B}_2 . Enfin, on référera aux éléments (de cadres différents) qui sont équivalents selon \sim par le même symbole ; on parlera, par exemple, d'un A qui appartient aussi bien à \mathbf{B}_1 qu'à \mathbf{B}_2 .

A.3.2 Règles comme opérations : Séquents et cadres_f

Dans un cadre_f donné, les symboles « logiques » qui apparaissent dans les règles **(Réf)**-**(Or)** ont une interprétation. L'implication matérielle \rightarrow correspond par exemple à l'ordre de l'algèbre interprétée (\leq), et la conséquence non-monotone \Rightarrow est « interprétée » selon la notion de satisfaction définie ci-dessus (Définition A.3.2). Alors, si le cadre_f remplit les conditions appropriées, les énoncés, ou pour le dire autrement, les « séquents », sont *admis* dans ce cadre_f. En l'occurrence, une implication \rightarrow est admis dans un cadre_f si l'antécédent et le conséquent entre dans l'ordre \leq approprié, et une conséquence \Rightarrow est admise si les conditions de la Définition A.3.2 sont satisfaites.

Il s'ensuit qu'un cadre_f valide ou admet certains énoncés ou séquents. Or, une règle permet le passage à ou l'admission d'autres énoncés ou séquents. Donc, une règle pourra être considérée comme permettant le passage à un cadre_f où ces autres énoncés ou séquents sont valides ou admises. C'est l'idée sous-jacente à la conception des règles logiques comme des contraintes sur des transitions ou des opérations sur les cadres.

La règle **(Réf)**, la règle la plus simple dans la mesure où elle ne comporte pas de prémisses et donc pas de « transformation » de cadres « anciens » dans un cadre « nouveau », servira d'illustration de l'idée générale.

²⁶On a discuté cette conséquence dans le cas de la structure instantanée de la conscience dans §§2.2.1 et 2.4.

²⁷Une relation de congruence est une relation d'équivalence qui respecte la structure algébrique sur les éléments, c'est-à-dire, en l'occurrence, les opérations booléennes sur les éléments des algèbres interprétées. Ces opérations peuvent être directement définies à partir des opérations sur les algèbres d'interprétation et les algèbres de base ; on fait cependant économie des détails, qui ne joueront pas un grand rôle dans ce qui suit.

(Réfl)

La règle **(Réfl)** a une conclusion, et aucune prémisse. Donc elle pose un cadre (validant la conclusion), sans avoir de cadre pré-donné (prémisse). Pour une conséquence $A \Rightarrow A$ posé par la règle **(Réfl)**, on peut choisir de prendre de cadre_f *minimal* validant cette conséquence, à savoir le cadre_f $(\mathbf{B}(\{\mathbf{A}\}, \emptyset), \prec)$ – le cadre_f engendré par l'énoncé A , sans autres contraintes (formules)²⁸, et avec comme ordre celui qui met $q(A) \prec q(\neg A)$. Ce cadre_f fait l'affaire quelle que soit la valuation, c'est-à-dire la valeur de vérité admise de $A : A \Rightarrow A$, selon l'Analyse **(ALP)**_r. En parenthèses, le choix d'ordre n'est pas important : l'ordre inverse aboutit au même résultat²⁹.

A.3.3 Règles comme opérations : Fusions, produits et quotients des algèbres

Pour les autres règles, il y a deux prémisses qui se « combinent » pour donner une conclusion. En termes de cadres, il s'agit de « fusionner » d'une manière ou d'une autre les cadres validant les prémisses pour donner un cadre où la conclusion est valide.

Pour ce faire, il s'agit de mobiliser des opérations mathématiques bien connues. D'abord on connaît l'opération de *produit libre*, qui se définit aussi bien sur les algèbres de Boole que sur les ordres : on s'en sert pour définir une notion de *produit libre*, d'abord sur les algèbres interprétées, et ensuite sur les cadres_f. Or, le produit libre ne prête pas d'attention aux possibles « identifications » entre les énoncés dans les différents cadres_f et algèbres interprétées : pour chaque paire d'énoncés, $A^1 \in \mathbf{B}_1$, $A^2 \in \mathbf{B}_2$, les images dans le produit libre des algèbres sont distinctes. Or, comme l'a été souligné fortement dans la considération des structures instantanées de la conscience (§2.4), la question de l'identification des aspects (en l'occurrence, des énoncés), entre les localités différentes (dans le cas présent, des cadres_f ou les algèbres interprétées) est aussi importante qu'elle est difficile³⁰. Dans §A.3.1, on a admis une relation de congruence \sim reliant

²⁸Rappeler de §5.1.1 qu'une algèbre interprétée pourrait être définie alternativement comme le quotient de l'algèbre d'interprétation par un certain ensemble d'énoncés. On utilise dans cette section $\mathbf{B}(I, \mathcal{F})$ pour l'algèbre interprétée avec algèbre interprétée B_I et algèbre de base $B_I/(\mathcal{F})$.

²⁹Pour le dire en d'autres termes, l'algèbre valide $A \Rightarrow A$ quelle que soit la « valuation » (le « monde » minimal_{\prec}).

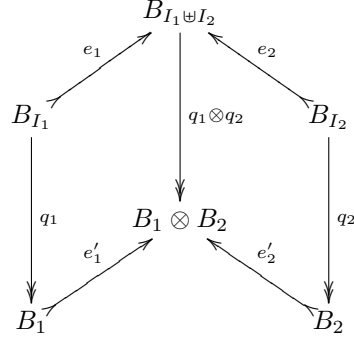
³⁰Alors que l'on n'a pas insisté sur ce point dans le cas des contrefactuels, il hérite la complexité de la question que l'on a mise en évidence dans le cas de la conscience est héritée, en raison du fait que la difficulté découle, pour l'essentiel, de la localité ou du *hors sujet*, qui joue un rôle important également dans la considération des contrefactuels (§2.2.1, §§4.2.4 et

des énoncés des algèbres interprétées différentes, et représentant l'identité entre les énoncés. Pour la bonne notion de « fusion » des cadres_f, et *a fortiori* des algèbres interprétées, il faut que les énoncés $A^1 \in \mathbf{B}_1$, $A^2 \in \mathbf{B}_2$ avec $A^1 \sim A^2$ ont la même image dans la « fusion », alors que, comme on vient de remarquer, leurs images sont distinctes dans le produit libre des algèbres. Pour ce faire, on dispose d'une opérations familière, à savoir le *quotient*. Le quotient d'une algèbre de Boole par une relation de congruence sur ces éléments rendra une nouvelle algèbre, avec une application prenant des éléments de l'ancienne dans la nouvelle, de telle sorte que les éléments équivalents sous la relation ont la même image. Pour ainsi dire, le quotient *identifie* les éléments qui était auparavant distincts. Il s'agit d'étendre cette opération, d'abord aux algèbres interprétées, ensuite aux cadres_f. La « fusion » de deux cadres_f (respectivement algèbres interprétées) pourra désormais être définie comme le produit libre suivie par le quotient sous la relation \sim qui représente l'identité des énoncés. Cette notion de « fusion », et plus généralement ces opérations sur les cadres_f, permettra de satisfaire des règles de la logique non monotone (§A.1), ainsi que de mieux comprendre leurs apparents cas d'échec.

On considère premièrement la question de la « fusion » de deux algèbres interprétées. L'opération de produit libre peut être définie sur les algèbres interprétées de manière suivante.

Définition A.3.3 (Produit libre des algèbres interprétées). Le *produit libre* de deux algèbres interprétées $\mathbf{B}_1 = (B_{I_1}, B_1, q_1)$ $\mathbf{B}_2 = (B_{I_2}, B_2, q_2)$ est l'algèbre interprétée $\mathbf{B}_1 \otimes \mathbf{B}_2 = (B_{I_1 \uplus I_2}, B_1 \otimes B_2, q_1 \otimes q_2)$, où \uplus est l'union disjoint (des ensembles) et \otimes est le produit libre (des algèbres de Boole, et par conséquent des applications entre algèbres de Boole).

Cette définition fait du sens – en particulier, l'application de quotient est le produit libre des applications de quotient – en raison du fait que $B_{I_1 \uplus I_2} = B_{I_1} \otimes B_{I_2}$. Alors, la situation est représentée par le diagramme suivant (les e sont les plongements canoniques des algèbres dans leurs produits).



Comme l'a été déjà remarqué, les (images des) éléments des algèbres de départ restent distinct(e)s dans le produit : si \mathbf{B}_1 et \mathbf{B}_2 comportent les éléments $A_{\mathbf{B}_1}$ et $A_{\mathbf{B}_2}$ respectivement, il y aura deux éléments différents – A^1 et A^2 , qui sont les images des A sous les plongements canoniques dans $\mathbf{B}_1 \otimes \mathbf{B}_2$ ³¹

Pour respecter l'identité des éléments, caractérisée par la relation \sim , on emploie l'opération algébrique de *quotient*, qui était déjà définie, sur les algèbres de Boole, dans la 5.1.1 (Définition 5.1.4). Pour « identifier » deux éléments x, y d'un algèbre de Boole B , on peut prendre le quotient par leur différence symétrique, $B/(x \triangle y)$ ³², qui est obtenue, rappelle-t-on, en « soustrayant » tous les éléments $z \leq x \triangle y$, c'est-à-dire tous les éléments qui « témoignent » la différence entre x et y . En telle sorte que, dans $B/(x \triangle y)$, l'image de x et l'image de y coïncident ; en ce sens ils sont identifiés.

Dans le cas des algèbres interprétées, il s'agit d'appliquer le quotient aussi bien à l'algèbre d'interprétation qu'à l'algèbre de base.

Définition A.3.4 (Quotient d'une algèbre interprétée). Soit \mathbf{B} une algèbre interprétée et A, B deux éléments. Le *quotient de \mathbf{B} par la différence entre A et B* , $\mathbf{B}/(A \sim B)$ est la structure \mathbf{B}' consistant en des algèbres $B_I/(A \triangle B)$, $B/(q(A) \triangle q(B))$ et l'application de quotient engendrée entre elles. Sous forme de diagramme³³ :

$$\mathbf{B}' \left\{ \begin{array}{ccc} B_I/(A \triangle B) & \xleftarrow{q_{A \triangle B}} & B_I \\ \downarrow q' & & \downarrow q \\ B/(q(A) \triangle q(B)) & \xleftarrow{q_{q(A) \triangle q(B)}} & B \end{array} \right\} \mathbf{B}$$

³¹Il convient de souligner que cette discussion porte sur les éléments des algèbres interprétées, c'est-à-dire des éléments des algèbres d'interprétations avec les éléments de l'algèbre de base qui leurs sont associés par l'application de quotient.

³² $x \triangle y = (x \vee y) \setminus (x \wedge y)$.

³³Rappel de la Définition 5.1.4 : l'application de B dans $B/(x \triangle y)$ est écrite $q_{x \triangle y}$ et appelée *application de quotient*.

On appelle l'application « de quotient » entre \mathbf{B} et \mathbf{B}' $q_{A\Delta B}$.

De manière générale, pour une relation de congruence \sim sur \mathbf{B} , \mathbf{B}/\sim est le résultat de quotienter par les différences symétriques entre tous A, B avec $A \sim B$, et q_\sim est l'application de quotient appropriée.³⁴

Remarque A.3.4. Selon la définition stricte d'algèbre interprétée (Définition 5.1.3), la structure qui vient d'être définie n'est pas nécessairement une algèbre interprétée, car $B_I/(A\Delta B)$ n'est pas nécessairement engendrée librement par un ensemble d'énoncés. Ce n'est dans des cas spéciaux (par exemple, où $A, B \in I$) que \mathbf{B}' est une algèbre interprétée; les opérations de rétrécissement et de restriction (Définitions 5.1.9 et 5.1.10 de §4.2.2) sont des exemples de tels cas spéciaux. Malgré cette subtilité, on continuera d'appeler l'algèbre \mathbf{B}' une algèbre interprétée, en sous-entendant une définition plus relâchée, mais non moins facile à préciser, d'algèbre interprétée.

Pour une paire d'algèbres interprétées, et une relation de congruence entre elles (reliant les « énoncés identiques »), on pourrait définir une notion de « fusion », $*$.

Définition A.3.5 (Fusion des algèbres interprétées). Pour $\mathbf{B}_1, \mathbf{B}_2$ deux algèbres interprétées avec une relation de congruence \sim sur l'union disjointe de leur éléments, avec $A \approx B$ pour $A \neq B$, $A, B \notin \mathbf{B}_i$ pour quelque i , alors la fusion

$$\mathbf{B}_1 * \mathbf{B}_2 = (\mathbf{B}_1 \otimes \mathbf{B}_2) / \sim$$

où \sim est l'image de la relation de congruence *sim* dans le produit libre.

On pourrait illustrer désormais ces opérations sur les règles **(ELG)** et **(AD)**, car, pour chacune d'elles, une des prémisses ne figure pas la conséquence non-monotone \Rightarrow , et donc l'ordre sur le cadre_f n'est pas importante.

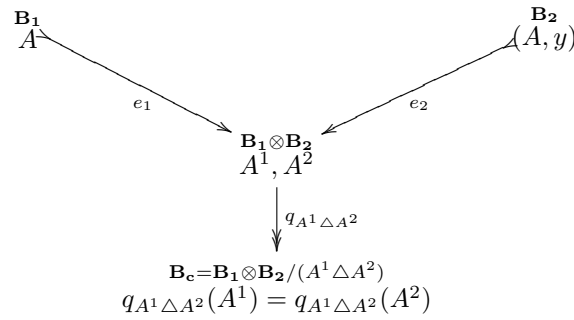
(ELG)

Les deux prémisses sont $A \leftrightarrow B$ et $A \Rightarrow C$. D'un côté, un cadre (\mathbf{B}_2, E_2) valide la deuxième prémisses si E_2 satisfait les conditions de l'Analyse **(ALP)**. De l'autre côté, la première prémisses correspond à une propriété structurelle du cadre_f approprié – la propriété que les deux éléments y sont équivalents dans

³⁴Ces notions sont bien-définies grâce à la commutation des quotients. Par ailleurs, cette définition se ramène, bien entendu, à la définition classique du quotient par une relation de congruence (l'algèbre des classes de congruence); si l'on présente le quotient à partir de la différence symétrique entre deux éléments, c'est pour faire le rapport avec la Définition 5.1.4 et l'idée du quotient comme la « soustraction » des éléments, ainsi que pour mettre l'accent sur « l'identification » des éléments.

son algèbre interprétée – qui est indépendant de l'ordre. \mathbf{B}_1 valide $A \leftrightarrow B$ s'il comporte A et B et si, dans \mathbf{B}_1 , $A \equiv B$; le cas le plus simple est l'algèbre $\mathbf{B}(\{A, B\}, \{A \leftrightarrow B\})$ ³⁵. (Pour fixer les idées, on peut fixer l'ordre comme $q(\neg A) \prec_2 q(A)$.)

L'application de la règle **(ELG)** « fusionne » en quelque sens ces deux algèbres interprétées (ou, à strictement parler, les cadres_f). Soit le cas où \sim n'identifie que les deux éléments A des deux algèbres interprétées qui figurent dans les prémisses. Sous forme de diagramme, la fusion se représente ainsi³⁶ :



Dans la fusion $\mathbf{B}_1 * \mathbf{B}_2 = \mathbf{B}_1 \otimes \mathbf{B}_2 / (A \triangle A)$ où les (images) des deux énoncés A des algèbres du départ coïncident. Puisque $A \equiv B$ dans \mathbf{B}_1 , et les opérations préservent cette équivalence, on a toujours la même équivalence dans $\mathbf{B}_1 * \mathbf{B}_2$. De plus, il est facile de « transférer » l'ordre \prec_2 sur l'algèbre \mathbf{B}_2 sur $\mathbf{B}_1 * \mathbf{B}_2$, pour obtenir un ordre tel que le cadre_f valide $B \Rightarrow C$ (l'ordre sur \mathbf{B}_1 , dont seulement les équivalences importent, ne joue aucun rôle). La composition d'un produit avec un quotient, c'est-à-dire la « fusion » appropriée des cadres correspondant aux prémisses, fournit (un cadre validant) la conclusion : elle « respecte » ou « retrouve » la règle **(ELG)**.

Proposition A.3.1 ((ELG)). *Si (\mathbf{B}_1, \prec_1) et (\mathbf{B}_2, \prec_2) satisfont les deux prémisses de la règle **(ELG)**, et \sim relie seulement les deux énoncés A en jeu dans les deux prémisses, alors le cadre_f $(\mathbf{B}_1 * \mathbf{B}_2, \prec'_2)$ valide la conclusion, où \prec'_2 est l'image de \prec_2 dans le quotient.*

Remarque A.3.5. Si l'énoncé B n'appartient pas à (\mathbf{B}_1, \prec_1) , alors la fusion correspond effectivement à une opération de distension (Définition 5.1.11) ; en fait, elle est la simple inverse d'un rétrécissement qui enlève l'énoncé B (équivalent à A dans l'algèbre) du cadre final (\mathbf{B}, E) . C'est une manière plus simple de comprendre la règle : elle ajoute l'énoncé B à la structure logique en vigueur, en tant qu'énoncé équivalent à A .

³⁵Pour des rappels terminologiques, voir la Définition 5.1.5 et notamment la note 28.

³⁶Voir les définitions pour la terminologie.

Ce résultat technique connaît cependant une limite *en pratique*, où \sim pourrait relier plusieurs énoncés entre les deux algèbres.

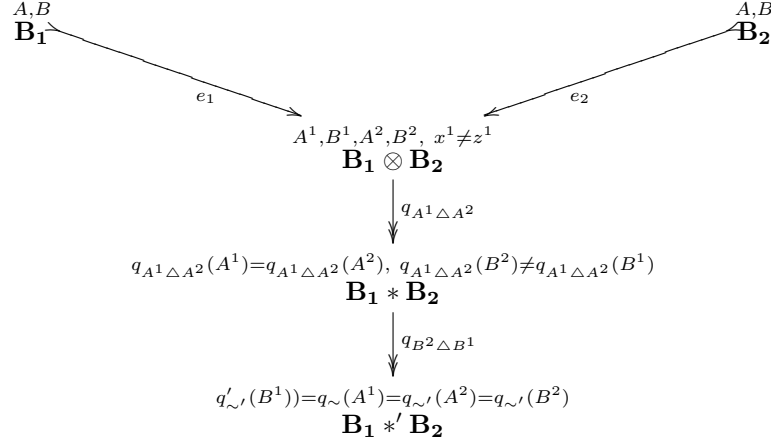
Le genre de complication est illustré par un cas où l'énoncé B apparaît déjà dans (\mathbf{B}_1, \prec_1) , ou, plus précisément par un cas où l'on emploie \sim' qui est comme \sim sauf qu'elle relie, en plus, l'élément B de \mathbf{B}_2 à un élément (B) de \mathbf{B}_1 ³⁷. Il est facile à voir que la fusion des algèbres utilisant cette relation de congruence, $\mathbf{B}_1 *' \mathbf{B}_2$ est obtenue à partir de la fusion utilisant l'ancienne relation, $\mathbf{B}_1 * \mathbf{B}_2$, avec une opération de quotient supplémentaire.

Par la Proposition A.3.1, la conséquence $B \Rightarrow C$ est validée dans $\mathbf{B}_1 * \mathbf{B}_2$, mais relative à B^2 , l'image du B de \mathbf{B}_2 . Or, puisque, dans $\mathbf{B}_1 * \mathbf{B}_2$, les images des éléments B sont éventuellement différents, cela n'implique pas que la conséquence est validée pour B^1 (l'image du B de \mathbf{B}_1). L'effet du quotient supplémentaire qui donne $\mathbf{B}_1 *' \mathbf{B}_2$, et par suite l'évaluation de $B \Rightarrow C$ dans $\mathbf{B}_1 *' \mathbf{B}_2$, dépend du rapport entre B^1 et B^2 dans $\mathbf{B}_1 * \mathbf{B}_2$. Il y a effectivement deux possibilités.

$A \leftrightarrow B$ **valide dans** (\mathbf{B}_1, \prec_1) Dans $\mathbf{B}_1 * \mathbf{B}_2$, on a déjà $A \equiv B^1 \equiv B^2$. Un autre quotient (des algèbres interprétées), « identifiant » les deux B , pour aboutir à $\mathbf{B}_1 *' \mathbf{B}_2$ (avec relation de congruence \sim'), ne change donc pas l'algèbre de base ; elle ne fait que « identifier » deux énoncés de l'algèbre d'interprétations qui sont déjà équivalents. En effet, $\mathbf{B}_1 *' \mathbf{B}_2$ est isomorphe à (\mathbf{B}_1, \prec_1) , ce qui est tout à fait naturel, étant donné que, puisque $A \equiv B$ dans (\mathbf{B}_1, \prec_1) , cette instance de la règle est déjà valide *localement*, *dans le cadre* (voir §A.2).

$A \leftrightarrow B$ **dans** (\mathbf{B}_1, \prec_1) Le quotient supplémentaire qui donne $\mathbf{B}_1 *' \mathbf{B}_2$ (celui qui identifie les deux B^i), on modifie la structure reliant les (images des) éléments de \mathbf{B}_1 . Sous forme de diagramme :

³⁷Plus précisément : \sim' est la relation de congruence minimale qui contient \sim et qui contient (B, B) .



Les images des A et B dans $\mathbf{B}_1 *' \mathbf{B}_2$ sont identiques, alors que les deux éléments sont distincts dans \mathbf{B}_1 . En ce sens, l'identification des deux énoncés B provenant des algèbres différentes *modifie* la structure héritée de l'algèbre de départ. Il se peut donc que l'image de l'ordre \prec'_2 perd certain de ses propriétés, que la conséquence $A \Rightarrow C$ (et *a fortiori* $B \Rightarrow C$) ne soit plus valide, et ainsi de suite. On aura l'occasion de revenir sur ce genre de complexité dans la discussion des autres règles.

(AD)

Comme (ELG) sauf qu'il faille veiller à ce que B soit dans B_E .

A.3.4 Règles comme opérations : Fusions, produits et quotients des cadres_f

Pour pouvoir rendre compte des autres règles non monotones (§A.1), et de manière générale les étapes d'inférence figurant plusieurs conséquences (\Rightarrow), il faut étendre les opérations définies ci-dessus sur les algèbres interprétées, ou les *structures logiques locales* si l'on veut, aux cadres ou cadres_f susceptibles à évaluer des conséquences (selon l'Analyse (ALP) ou (ALP)_r par exemple). Puisqu'il s'agit en l'occurrence des cadres_f, c'est-à-dire des algèbres interprétées munies des ordres particuliers (§A.3.1), il s'agit de définir des opérations sur ces ordres susceptible de les « fusionner ». Comme dans le paragraphe précédent, on procède en trois temps. Le premier consiste à définir un produit sur les ordres qui va de paire avec le produit libre sur les algèbres, et donc un produit des cadres_f. Ensuite, on considère le comportement des ordres et par conséquent des cadres_f sous l'opération de quotient appliquées aux algèbres interprétées,

pour aboutir enfin à une définition de « fusion » de cadres_f comme la composition d'un produit avec des quotients. Cette opération de fusion satisfait aux dernières règles non monotones ((**Coup**) - (**Or**)) et permet en outre de comprendre leurs apparents contre-exemples.

Pour commencer avec la définition d'une opération de produit sur les cadres_f, c'est-à-dire sur les algèbres interprétées munies des ordres satisfaisant des propriétés particulières, il s'agit de mobiliser l'opération mathématique bien connue du produit des *ordres*.

Définition A.3.6 (Produit des ordres). Soient \prec_1 et \prec_2 deux ordres sur deux ensembles S_1 et S_2 . L'ordre produit $\prec = \prec_1 \times \prec_2$ est l'ordre sur $S_1 \times S_2$ défini par :

$$(a, b) \prec (c, d) \text{ ssi } \begin{cases} a \prec_1 b \ \& \ c = d \\ a = b \ \& \ c \prec_2 d \\ a \prec_1 b \ \& \ c \prec_2 d \end{cases}$$

La pertinence de cette opération est conséquence au fait suivant, qui assure que ce produit préserve les propriétés des ordres des cadres_f.

Fait A.3.1. *Si \prec_1 et \prec_2 sont irréflexifs, transitifs et lisses, $\prec_1 \times \prec_2$ l'est aussi.*

En effet, si S_1 et S_2 sont des ensembles des éléments des algèbres interprétées \mathbf{B}_1 et \mathbf{B}_2 qui satisfont aux conditions de la Définition A.3.1, alors $\{s^1 \wedge t^2 \mid s \in S_1, t \in S_2\}$ est un ensemble d'éléments de $\mathbf{B}_1 \otimes \mathbf{B}_2$ qui satisfait également les conditions de la Définition A.3.1. De plus, cet ensemble est isomorphe au produit cartésien $S_1 \times S_2$. Par conséquent, l'opération suivante de produit des cadres_f est bien définie.

Définition A.3.7 (Produit des cadres_f). Le produit de deux cadres_f (\mathbf{B}_1, \prec_1) et (\mathbf{B}_2, \prec_2) , $(\mathbf{B}_1, \prec_1) \otimes (\mathbf{B}_2, \prec_2)$ est le cadre_f ayant

- pour algèbre interprétée $\mathbf{B}_1 \otimes \mathbf{B}_2$;
- pour ordre $\prec_1 \times \prec_2$.

Comme discuté ci-dessus, il ne suffit pas de prendre le produit libre des algèbres interprétées, car il ne respecte pas l'*identité* des énoncés figurant dans des algèbres différents, laquelle est représentée par la relation de congruence \sim . Pour ce faire, il fallait prendre des quotients des algèbres interprétées ; il s'agit maintenant d'examiner l'effet de l'opération de quotient sur l'ordre (produit) dans le produit libre des deux algèbres, pour, par la suite, pouvoir définir une notion de « fusion » entre les cadres_f qui respecte l'identité des énoncés, c'est-à-dire, qui respecte la relation \sim .

Puisqu'il s'agira pour l'essentiel d'identifications des énoncés qui sont impliqués dans des conséquences particulières, et donc aux énoncés dans B_S^{38} , il suffit de considérer le cas des quotients qui identifient des éléments des B_S . Dans de tels cas, le fait suivant est capital.

Fait A.3.2. *Soit S une partition d'une algèbre de Boole B , $x \in B_S$ et $q_x : B \rightarrow B/(x)$ l'application de quotient. Alors*

$$\begin{aligned} \forall a \in S, \quad & q_x(a) = \perp \text{ ssi } a \leq x \\ \forall a, b \in S \text{ tq } q_x(a), q_x(b) \neq \perp \quad & q_x(a) \wedge q_x(b) = \perp, \\ \text{et} \quad & \bigvee_{\substack{a_i \in S \\ q_x(a_i) \neq \perp}} q_x(a_i) = \top, \\ \text{et} \quad & q_x|_S : S \rightarrow \hat{q}_x(S) \text{ est injective} \end{aligned}$$

où $\hat{q}_x(S) = \{q_x(a) \mid a \in S, a \not\leq x\}$

En conséquence, l'image de S sous q_x est une partition de $B/(x)$ (au sens de la Remarque A.3.1) ; de plus, q_x est une bijection sur les éléments de S qui ne sont pas l'image de \perp . Intuitivement, le quotient ne fait qu'« soustraire » de S tous les éléments qui sont « au-dessous » de x ; ce faisant, il ne touche pas aux propriétés qui le font une partition.

De plus,

Fait A.3.3. *Soit \prec un ordre irréflexif, transitif et lisse sur un ensemble S , et S' un sous-ensemble de S . Alors, \prec' , la restriction de \prec à S'^{39} , est irréflexif, transitif et lisse.⁴⁰*

Il s'ensuit un comportement simple de l'ordre sur S sous un quotient de l'algèbre (par un élément de S) : notamment, la *restriction* de l'ordre d'un cadre _{f} à l'image de son domaine sous l'application de quotient a toutes les propriétés d'un ordre d'un cadre _{f} . D'où l'opération suivante est une opération bien définie sur les cadres _{f} .

Définition A.3.8 (Quotient des cadres _{f}). Soit (\mathbf{B}, \prec) un cadre _{f} et A, B deux éléments dans B_S^{41} . Le *quotient* de (\mathbf{B}, \prec) par la différence entre A et B , a

– pour algèbre interprétée $\mathbf{B}' = \mathbf{B}/(A \sim B)$;

³⁸Voir les Définitions A.3.1 et A.2.4 pour les restrictions sur les conséquences qui peuvent être évaluées, et les conventions terminologiques introduites dans §5.1.2.

³⁹ $\forall x, y \in S', x \prec' y \text{ ssi } x \prec y$.

⁴⁰Voir les notes 8 et 18 sur le caractère trivial de la propriété d'être lisse dans les cas finis.

⁴¹On garde la terminologie de la Définition A.3.1, selon laquelle S est l'ensemble où s'applique \prec .

– pour ordre \prec' , la restriction de \prec à $\widehat{q}_{A\Delta B}(S)$ ⁴².

De manière générale, pour une relation de congruence \sim sur les éléments de (\mathbf{B}, \prec) , $(\mathbf{B}, \prec)/\sim$ est le résultat de quotienter par les différences entre A et B pour tous A, B tels que $A \sim B$.

L'opération de quotient rend un cadre_f, mais il n'est pas nécessaire que ce cadre_f donne la même évaluation aux conséquences que le cadre_f original. Autrement dit, le fait de restreindre l'ordre à un sous-ensemble peut changer l'évaluation de conséquences, comme témoigne le cadre avec les trois éléments suivants dans l'ordre indiqué⁴³, et le quotient par $A \Delta B$. Avant le quotient, qui implique le soustraction du second élément, $A \Rightarrow C$ est valide, alors qu'après, elle ne l'est pas.

$$\begin{array}{ccc} s_1 & & s_2 & & s_3 \\ \neg A, \neg B, C & \prec & A, \neg B, C & \prec & A, B, \neg C \end{array}$$

Remarque A.3.6. Le rapport entre les relations de conséquence avant et après le quotient n'est pas difficile à préciser, si l'on recourt au fait qu'une algèbre interprétée pourrait être considérée comme le quotient de l'algèbre d'interprétation par certains énoncés, qui sont, dans l'algèbre interprétée obtenue, équivalente au vrai (voir les remarques dans §5.1.1, notamment celles qui suivent les Définitions 5.1.3 et 5.1.4). Les seules complexités tiennent aux détails de la formulation ; en en faisant l'économie, il s'avère que, en gros, (\mathbf{B}', \prec') , le quotient de (\mathbf{B}, \prec) par la différence entre A et B , valide $C \Rightarrow D$ si et seulement si (\mathbf{B}, \prec) valide $C \& (A \leftrightarrow B) \Rightarrow D$.

Définition A.3.9 (Fusion des cadres_f). Pour (\mathbf{B}_1, \prec_1) et (\mathbf{B}_2, \prec_2) deux cadres_f avec une relation de congruence \sim sur l'union disjointe de leur éléments, satisfaisant aux conditions spécifiées dans la Définition A.3.5, alors la fusion

$$(\mathbf{B}_1, \prec_1) * (\mathbf{B}_2, \prec_2) = ((\mathbf{B}_1, \prec_1) \otimes (\mathbf{B}_2, \prec_2)) / \sim$$

où \sim est l'image de la relation de congruence \sim dans le produit libre.

Cette opération permet une analyse approfondie des règles qui restent de la logique non monotone.

⁴² $q_{A\Delta B}(C) \prec' q_{A\Delta B}(D)$ ssi $C \prec D$.

⁴³Les énoncés indiqués sont ceux dans B_S qui sont \geq ces éléments, ou autrement ceux qui sont « vrais » dans ces « états ».

(Coup)

Soit (\mathbf{B}_1, \prec_1) un cadre_f où la première prémisse $(A \wedge B \Rightarrow C)$ est valide, et (\mathbf{B}_2, \prec_2) un cadre_f où la deuxième prémisse $(A \Rightarrow B)$ est valide. Soit $(A \wedge B)^1$, C^1 , A^2 et B^2 les images des éléments appropriés dans le produit libre de ces cadres_f, $(\mathbf{B}_1 \otimes \mathbf{B}_2, \prec_1 \times \prec_2)$. Évidemment, $A^2 \Rightarrow C^1$ n'est pas nécessairement satisfaite dans le produit : puisqu'un élément $\text{minimal}_{\prec} A^2$ est la conjonction de l'image d'un élément $\text{minimal}_{\prec_2} A^2$ et l'image d'un élément minimal_{\prec_1} , cette conséquence est satisfaite seulement si C est impliquée par tous les éléments minimaux _{\prec_1} ⁴⁴. Il faut prendre en compte l'*identité* entre énoncés, ou, autrement dit, respecter la relation \sim entre les cadres_f ; ou, autrement dit, prendre la fusion $*$ par rapport à cette relation de congruence.

Pour n'importe quelle paire des algèbres qui valident les prémisses, le seul énoncé qu'il partage (et qui est dans le B_S) nécessairement est $A \wedge B$ ⁴⁵. On considère d'abord la relation de congruence « minimale » entre les cadres_f, à savoir la relation qui relie seulement le $A \wedge B$ d'un des cadres au $A \wedge B$ de l'autre⁴⁶ ; on appelle cette relation \sim .

Il n'est pas difficile de montrer la proposition suivante⁴⁷.

Proposition A.3.2 ((Coup)). *Pour deux cadres_f (\mathbf{B}_1, \prec_1) , (\mathbf{B}_2, \prec_2) dont la première valide $A \wedge B \Rightarrow C$ et la deuxième $A \Rightarrow B$, et \sim reliant seulement les deux énoncés $A \wedge B$ des deux cadres_f⁴⁸, alors la fusion $(\mathbf{B}_1, \prec_1) * (\mathbf{B}_2, \prec_2)$ satisfait $A^2 \Rightarrow C^1$. En ce sens, cette opération de fusion satisfait la règle (Coup).*

Or, de même que dans le cas de la règle (ELG) discutée ci-dessus, la situation devient plus complexe s'il y a plusieurs identifications des énoncés entre les cadres_f. Dans la discussion précédente (§A.3.3), on a vu que, à la différence du cas où la relation de congruence ne relie qu'une seule paire d'énoncés, la

⁴⁴La notion et terminologie $\text{minimal}_{\prec} A$, introduite lors de la Définition A.1.1, s'applique également ici : c'est l'élément minimal selon \prec où A .

⁴⁵Ceci est conséquence d'un côté du fait que B_S est clos par combinaison logique et de l'autre côté du fait qu'une algèbre interprétée pourrait contenir l'énoncé $A \wedge B$ sans contenir A ou B (si $A \wedge B \in I$, $A, B \notin I$, par exemple ; voir la Définition 5.1.3).

⁴⁶Plus précisément : la relation de congruence minimale qui contient $(A \wedge B, A \wedge B)$. Plutôt que d'utiliser cette formulation précise, on a choisi, afin de simplifier un peu l'exposition, d'utiliser l'expression relâchée dans la discussion qui suit.

⁴⁷Comme ailleurs dans ce travail, les détails des preuves sont omis. Dans le cas actuel, le résultat repose sur le fait que l'identification des deux $A \wedge B$ assure que les éléments minimaux _{\prec_c} A^2 impliquent également $(A \wedge B)^1$, de telle sorte que ces éléments ont pour « facteur provenant de \mathbf{B}_1 » non pas les éléments minimaux _{\prec_1} mais les éléments minimaux _{\prec_1} $A \wedge B$.

⁴⁸Voir la note 46.

fusion avec une relation de congruence qui relie plusieurs paires d'énoncés peut « changer » la structure de l'algèbre interprétée. De manière semblable, la fusion de deux cadres_f avec une relation reliant une paire d'énoncés seulement ne touche pas à la « structure logique », dans la mesure où la fusion valide non seulement les mêmes implications (\rightarrow) mais également les mêmes conséquences (\Rightarrow) que les cadres_f du départ. Ce qui n'est pas nécessairement le cas si l'on identifie plusieurs paires d'énoncés.

Pour exprimer ce premier aspect, il convient de définir la notion de préservation de conséquences.

Définition A.3.10 (Préservation des conséquences). Soient deux cadres_f (\mathbf{B}_1, \prec_1) et (\mathbf{B}_2, \prec_2) , et un cadre_f (\mathbf{B}, \prec) , avec des applications des cadres de (\mathbf{B}_i, \prec_i) dans (\mathbf{B}, \prec) . On dit que (\mathbf{B}, \prec) *préserve les conséquences* des deux autres cadres si, pour toute paire d'éléments D, E de (\mathbf{B}_i, \prec_i) , (\mathbf{B}_i, \prec_i) valide $D \Rightarrow E$ si et seulement si (\mathbf{B}, \prec) valide $D^i \Rightarrow E^i$ (où D^i est l'image de D dans (\mathbf{B}, \prec)).

Que la fusion de deux cadres_f avec une relation de congruence qui ne relie qu'une paire d'énoncés entre eux préserve les conséquences découle de l'existence des applications appropriées entre les cadres_f. Pour préciser ces applications, on a besoin de la définition suivante.

Définition A.3.11. Une application entre deux algèbres interprétées $\psi : \mathbf{B} \rightarrow \mathbf{B}'$ est la donnée de deux homomorphismes d'algèbres de Boole, $\psi_I : B_I \rightarrow B'_I$ et $\psi_B : B \rightarrow B'$ tel que le diagramme suivant commute⁴⁹ :

$$\mathbf{B}' \left\{ \begin{array}{ccc} B_I & \xrightarrow{\psi_I} & B'_I \\ \downarrow q & & \downarrow q' \\ B & \xrightarrow{\psi_B} & B' \end{array} \right\} \mathbf{B}$$

Une application ψ est *injective* si ψ_I et ψ_B sont injectives.

On peut maintenant définir la relation suivante entre cadres, et ensuite entre cadres_f.

Définition A.3.12. Pour deux cadres $(\mathbf{B}, E)_1$ et $(\mathbf{B}, E)_2$, $(\mathbf{B}, E)_1 \preceq (\mathbf{B}, E)_2$ si et seulement si

- il existe une application injective $\phi : \mathbf{B}_1 \rightarrow \mathbf{B}_2$ (Définition A.3.11) ;
- $\{\phi(A) \mid A \in E_1\} \subseteq E_2$.

⁴⁹On emploie toujours la terminologie de la Définition 5.1.3, selon laquelle B_I est l'algèbre d'interprétation et B l'algèbre de base.

De manière semblable, pour deux cadres_f $(\mathbf{B}, \prec)_1$ et $(\mathbf{B}, \prec)_2$, $(\mathbf{B}, \prec)_1 \preceq (\mathbf{B}, \prec)_2$ si et seulement si

- il existe une application injective $\phi : \mathbf{B}_1 \rightarrow \mathbf{B}_2$;
- $\forall a \in S_1, \phi(a) \in B_{S_2}$;
- pour tout $a, b \in S_1$, $\phi(a) \prec_2 \phi(b)$ ssi $a \prec_1 b$, où l'ordre \prec_2 est étendu aux éléments de B_{S_2} ⁵⁰ comme suit : $x \prec_2 y$ ssi pour tout $c \in S_2$, $c \leq x$, il n'existe pas de $d \in S_2$, $d \leq y$, $d \prec_2 c$.

Remarque A.3.7. Deux cadres entrent dans l'ordre \preceq s'il y a un plongement de l'algèbre interprétée de l'un dans celle de l'autre, qui prend les éléments de l'ensemble co-indépendant aux éléments de l'ensemble co-indépendants. Il n'est pas difficile à voir que la relation sur les cadres_f n'est qu'une extension de cette idée aux cadres_f. De fait, si $(\mathbf{B}, E)_1 \preceq (\mathbf{B}, E)_2$ en tant que cadres, alors $(\mathbf{B}, E)_1 \preceq (\mathbf{B}, E)_2$ en tant que cadres_f.

L'intérêt de cette relation tient à la propriété suivante.

Proposition A.3.3. *Soit $(\mathbf{B}, E)_1 \preceq (\mathbf{B}, E)_2$, avec application ϕ entre les deux algèbres interprétées. Alors, $\forall A, B \in \mathbf{B}_1$,*

$$(\mathbf{B}, E)_1 \text{ satisfait } A \Rightarrow B \text{ ssi } (\mathbf{B}, E)_2 \text{ satisfait } \phi(A) \Rightarrow \phi(B)$$

De même pour les cadres_f.

Or, en revenant à la question des rapports entre les cadres_f et leurs fusions, il s'avère y avoir ce genre de relation entre eux.

Fait A.3.4. *Pour deux cadres_f (\mathbf{B}_1, \prec_1) , (\mathbf{B}_2, \prec_2) , et une relation \sim qui identifie seulement les énoncés $A \in B_{S_1}$, $A \in B_{S_2}$, pour chaque i , $(\mathbf{B}_2, \prec_2) \preceq ((\mathbf{B}_1, \prec_1) \otimes (\mathbf{B}_2, \prec_2)) / \sim$.*

Il s'ensuit de la Proposition A.3.3 que la fusion avec ce genre de relation de congruence \sim préserve les conséquences.

Corollaire A.3.1. *Pour deux cadres_f (\mathbf{B}_1, \prec_1) , (\mathbf{B}_2, \prec_2) , et une relation \sim qui identifie seulement les énoncés $A \in B_{S_1}$, $A \in B_{S_2}$, $((\mathbf{B}_1, \prec_1) \otimes (\mathbf{B}_2, \prec_2)) / \sim$ en préserve les conséquences.*

En revanche, pour des fusions avec des relations reliant plus d'une paire d'énoncés, la situation est plus complexe. Comme le montrent les deux exemples suivants, il y aura des cas où la préservation des conséquences et la validité de la règle entrent en conflit : il n'est pas possible d'avoir les deux. Souvent, on

⁵⁰On rappelle qu'elle a été définie originellement sur les éléments de S_2 .

perd la préservation des conséquences. Or, alors que dans certains cas (Exemple A.3.1), cela suffit pour garder la validité de la règle, dans d'autres cas (Exemple A.3.2), la validité de la règle est également perdue.

Exemple A.3.1. Soit (\mathbf{B}_1, \prec_1) le cadre_f qui satisfait la première prémisse de **(Coup)** dont l'ordre sur les éléments pertinents prend cette forme :

$$\begin{array}{ccc} s_1^1 & & s_2^1 & & s_3^1 \\ \neg A, \neg B, C & \prec & A, \neg B, \neg C & \prec & A, B, C \quad \dots \end{array}$$

et (\mathbf{B}_2, \prec_2) , le cadre_f qui satisfait la deuxième prémisse, en comportant deux « états » dans l'ordre suivant :

$$\begin{array}{ccc} s_1^2 & & s_2^2 \\ \neg A, \neg B & \prec & A, B \end{array}$$

Pour la relation de congruence qui relie seulement les $A \wedge B$, $(\mathbf{B}_1, \prec_1) * (\mathbf{B}_2, \prec_2)$ valide $A^2 \Rightarrow C^1$ et $A^1 \Rightarrow \neg C^1$. Ce premier fait est d'ailleurs conséquence de la Proposition A.3.2, alors que le dernier est résumé dans le Corollaire A.3.1 (puisque (\mathbf{B}_1, \prec_1) valide $A \Rightarrow C$). Cette situation, d'apparence étrange, est parfaitement cohérente, dans la mesure où les images des énoncés A des deux cadres_f (A^1 et A^2) sont distincts dans la fusion avec \sim .

Si, en revanche, on « identifie » les deux énoncés A des deux cadres_f, il est évidemment *impossible* de préserver *aussi bien* la validité de la règle du coupure (Proposition A.3.2) et la validité des conséquences dans les cadres_f du départ (Corollaire A.3.1). Il s'agit d'employer, au lieu de \sim , la relation \sim' qui relie de surcroît les énoncés A dans les deux cadres_f, on obtient $(\mathbf{B}_1, \prec_1) *' (\mathbf{B}_2, \prec_2)$ qui comporte un seul élément A , l'image des deux A de (\mathbf{B}_1, \prec_1) et (\mathbf{B}_2, \prec_2) , qui valide $A \Rightarrow C^{51}$. La conclusion de la règle **(Coup)** est validée, mais la conséquence $A \Rightarrow \neg C$, valide dans (\mathbf{B}_1, \prec_1) , n'est pas préservée.

Bref, si la relation de congruence \sim relie plus d'une paire d'énoncés, les conséquences ne sont pas nécessairement préservées dans la fusion (avec cette relation).

Pourtant, non seulement les conséquences ne sont pas préservées dans le cas d'identifications multiples, mais le « changement » de structure peut entraîner également une enfreinte de la règle **(Coup)**, comme le montre l'exemple suivant.

Exemple A.3.2. Soit (\mathbf{B}_1, \prec_1) le cadre_f de l'Exemple A.3.1 et (\mathbf{B}_3, \prec_3) , le cadre_f qui satisfait la deuxième prémisse, en comportant ces trois « états » dans l'ordre suivant :

⁵¹Pour ainsi dire, $s_1^1 \otimes s_2^2$ et $s_2^1 \otimes s_1^2$ sont soustraits dans le quotient, de sorte que l'élément le plus proche où A est $s_2 \otimes s_3$.

$$\begin{array}{ccc} s_1^3 & & s_2^3 \\ \neg A, \neg B & \prec & A, B \\ & & \prec \\ & & s_2^3 \\ & & A, \neg B \end{array}$$

De même que dans l'exemple précédent, pour la relation de congruence qui relie seulement les $A \wedge B$, $(\mathbf{B}_1, \prec_1) * (\mathbf{B}_3, \prec_3)$ valide $A^3 \Rightarrow C^1$ et $A^1 \Rightarrow \neg C^1$. Pourtant, si l'on identifie les A , en employant la relation \sim'' qui relie de surcroît les énoncés A dans les deux cadres_f, le cadre_f obtenu, $(\mathbf{B}_1, \prec_1) *'' (\mathbf{B}_2, \prec_2)$, ne valide ni $A \Rightarrow C$, ni $A \Rightarrow \neg C$ ⁵². Dans ce cas, donc, non seulement les conséquences ne sont pas préservées, mais la conclusion de la règle **(Coup)** n'est pas validée.

En résumé, si la relation de congruence \sim relie plus d'une paire d'énoncés, la règle **(Coup)** n'est pas nécessairement validée par la fusion $*$ avec \sim .

Ces exemples servent à mettre en évidence la spécificité de la fusion $*$ avec une relation qui identifie une seule paire d'énoncés : elle satisfait à la règle **(Coup)** (Proposition A.3.2) et elle préserve les conséquences (Corollaire A.3.1). Pour des relations reliant plusieurs paires d'énoncés, les conséquences ne sont pas nécessairement préservées, et la règle n'est pas nécessairement satisfaite.

En outre, ces observations suffisent pour expliquer les apparents contre-exemples à la règle **(Coup)**, tel celui qui a été présenté dans §4.3.1 et repris au début de la présente section.

Ce contre-exemple prend comme prémisses « S'il pleuvait (A), j'apporterais un parapluie (B) » et « s'il pleuvait (A) et j'apportais un parapluie (B), je ne me mouillerais pas (C) », et conclusion « s'il pleuvait (A), je ne me mouillerais pas (C) ». Un bon modèle pour la première prémisse aura, comme « état » minimal où il pleut (A), un « état » où j'apporte un parapluie (B); en ce sens, les cadres_f (\mathbf{B}_2, \prec_2) et (\mathbf{B}_3, \prec_3) des Exemples A.3.1 et A.3.2 servent de modèles. Or, puisque, dans (\mathbf{B}_2, \prec_2) , la conjonction du fait de pleuvoir et du fait que j'apporte un parapluie est (localement) « contradictoire » (il n'a pas d'« état » où $A \wedge \neg B$), alors qu'elle ne l'est pas selon (\mathbf{B}_3, \prec_3) , ce dernier constitue un modèle plus exact de la première prémisse⁵³. D'un autre côté, (\mathbf{B}_1, \prec_1) semble la façon la plus naturelle de représenter la deuxième prémisse : l'« état » minimal où il pleut et j'apporte un parapluie est un « état » où je ne me mouille pas,

⁵²Les éléments $s_2^1 \otimes s_3^3$ et $s_3^1 \otimes s_2^3$ sont les deux minimaux, mais ils diffèrent sur la valeur de C , pour ainsi dire.

⁵³On pourrait refaire l'analyse qui suit avec le cadre_f (\mathbf{B}_2, \prec_2) , et donc s'appuyant sur l'Exemple A.3.1. On aboutit à une explication du contre-exemple en termes de la tension entre les conséquences valides dans des cadres_f originels et les conséquences valides dans le cadre_f d'arrivée, tension qui existe parce qu'il y a des conséquences (tel $A \Rightarrow \neg C$) dont la préservation dans le cadre final entre en contradiction avec la validité de la règle.

alors qu'il pourrait y avoir un état inférieur où il pleut, je n'ai pas de parapluie et je me mouille. Ce qui n'est pas à dire que ce cadre, ou un cadre suffisamment semblable, est la seule possibilité, mais seulement qu'elle est la plus naturelle. Des considérations griciennes viennent à l'appui de la naturalité de ce cadre : elle exigent qu'une contribution soit aussi informative que nécessaire⁵⁴, de telle manière que l'« état intermédiaire » où il pleut, je n'ai pas mon parapluie et je me mouille est « suggéré » par l'usage de $A \& B \Rightarrow C$ plutôt que $A \Rightarrow C$ – pour employer le terme gricien, il en est une *implicature*. Or, si l'on admet les deux cadres_f (\mathbf{B}_1, \prec_1) et (\mathbf{B}_3, \prec_3) comme pertinents pour représenter ces deux contrefactuels, et si l'on admet, comme l'est naturel, d'identifier les énoncés « il pleut » (A) qui figurent dans les deux cadres, l'étrangeté de la conclusion de l'application de la règle **(Coup)** s'explique en termes des complexités mises en évidence dans l'Exemple A.3.2. À savoir, si l'on identifie les deux énoncés A (« il pleut ») entre ces deux cadres_f, la fusion $*$ ne valide pas la conclusion de la règle **(Coup)**. La règle n'est donc pas validée dans l'exemple, comme l'atteste le sens commun, largement en raison des multiples identifications, naturelles, entre les deux cadres_f qui valident les prémisses. Pour garder la validité la règle, il faut renoncer à identifier les deux énoncés A (« il pleut »), et affirmer que c'est le « il pleut » de la première prémissse qui figure dans la conséquence « s'il pleut, je ne me mouille pas », mais non pas le « il pleut » de la deuxième prémissse, qui, provenant d'un contexte différent, a un « sens différent ».

En résumé, la règle **(Coup)** est validée par l'opération de fusion $*$ dans le cas simple de l'identification des énoncés $A \wedge B$ seulement entre les deux cadres_f qui valident les prémisses. Or, dans des cas des identifications supplémentaires, elle risque de ne pas être satisfaite. Des résultats apparentés s'appliquent à d'autres règles, ainsi que les analyses semblables des contre-exemples apparents.

(MP)

Le résultat important est le suivant.

Proposition A.3.4 ((MP)). *Pour deux cadres_f (\mathbf{B}_1, \prec_1) , (\mathbf{B}_2, \prec_2) dont le premier valide $A \Rightarrow B$ et la deuxième $A \Rightarrow C$, et \sim la relation de congruence qui relie seulement les énoncés A des deux cadres_f, la fusion des cadres_f avec cette relation, $(\mathbf{B}_1, \prec_1) * (\mathbf{B}_2, \prec_2)$ valide $A \wedge B^1 \Rightarrow C^2$.*

Cette fusion caractérise donc la règle **(MP)**.

De plus, on peut rendre compte du contre-exemple à la règle **(MP)** mentionné dans §4.3.1 et au début de cette section en employant des considérations

⁵⁴Grice (1989, Ch. 2).

analogues à celles qui ont servi dans l'analyse du contre-exemple à **(Coup)**.

(Or)

Proposition A.3.5 ((Or)). *Pour deux cadres_f (\mathbf{B}_1, \prec_1) , (\mathbf{B}_2, \prec_2) dont le premier valide $A \Rightarrow C$ et la deuxième $B \Rightarrow C$, et \sim la relation de congruence qui relie les énoncés C des deux cadres_f, la fusion des cadres_f avec cette relation, $(\mathbf{B}_1, \prec_1) * (\mathbf{B}_2, \prec_2)$ valide $A^1 \vee B^2 \Rightarrow C$.*

Cette fusion caractérise donc la règle **(Or)**.

Il s'ensuit des considérations des trois derniers paragraphes (notamment des Propositions A.3.1, A.3.2, A.3.4 et A.3.5) que l'opération de fusion permet non seulement une « représentation » les étapes d'inférence de telle manière qu'elles valident les règles de base de la logique non monotone (§A.1), mais également une explication des apparents contre-exemples à ces règles.

A.3.5 Les règles sur les cadres

La discussion précédente des opérations qui correspondent aux règles logiques s'est déroulée en termes des *cadres_f* ; mais quelles conclusions s'appliquent à l'analyse proposée de la conséquence, l'Analyse **(ALP)**, qui s'appuie sur des *cadres* figurant des ensembles co-indépendants, ou, de manière équivalente, les modèles-**(ALP)** (§A.2.2) ?

On a déjà remarqué dans §A.3.1 que les conclusions et les techniques de §§A.3.2 et A.3.3 s'appliquent également aux cadres ; il reste donc à déterminer dans quelle mesure les considérations de §A.3.4 peuvent être étendues pour s'appliquer aux cadres.

Pour répondre à cette question, il importe de rappeler que les cadres_f sont plus faibles que les cadres, dans la mesure où ils ne satisfont pas nécessairement, premièrement, à la condition sur l'ensemble S où s'applique l'ordre (troisième condition de la Définition A.2.2) et, deuxièmement, aux deux conditions sur l'ordre lui-même (qu'il est structuré de treillis et qu'il respecte la largeur, quatrième clause de la Définition A.2.2 ; Remarque A.3.2). Inversement, les cadres sont des cadres_f qui possèdent ces propriétés supplémentaires.

Or, il s'avère que ces propriétés *ne sont pas préservés sous l'opération de fusion*. En d'autres termes, l'opération de fusion n'est pas une opération qui, donné deux cadres, fournit un *cadre*. En l'occurrence, il n'y a pas de problème avec le produit libre.

Fait A.3.5. Soit \prec_1, \prec_2 deux ordres irréflexifs sur deux ensembles S_1, S_2 avec, pour $i = 1, 2$,

1. $\exists x_i \in N \ |\{S_i = 2_i^x\}|$;
2. (**structuré de treillis**) la clôture réflexive de \prec_i est un treillis, dont l'élément minimal est appelé v_i . On définit le rang d'un élément $r_i(s)$ comme le n minimal tel qu'il y a une suite $v_i \prec_i s_1 \prec_i \cdots \prec_i s_n = s$;
3. (**respectant du largeur**) $|\{u \mid r_i(u) = n\}| = \binom{x_i}{n}$.

Alors, le produit libre des deux ordres $\prec_1 \times \prec_2$, sur l'ensemble $S_1 \otimes S_2$, satisfait aux conditions 1., 2. et 3.

En revanche, le quotient ne préserve pas ces propriétés.

Fait A.3.6. Soit \prec un ordre irréflexif sur un ensemble S , satisfaisant aux conditions 1., 2. et 3. ci-dessus. Alors, un ordre quotient $\prec_c = \prec|_{S_c}$ sur $S_c = S \setminus X$, $X \subset S$, $X \neq \emptyset$ est un ordre irréflexif qui ne satisfait pas nécessairement aux conditions 1., 2. et 3.

La préservation de l'irréflexivité a été déjà constatée (fait A.3.3). Il est facile de construire des exemples de non préservation des autres propriétés.

Exemple A.3.3. 1. Prend $S = \{s_1, s_2, s_3, s_4\}$, $S_c = \{s_1, s_2, s_3\}$;
 2. Prend \prec sur S tel que le supremum de s_1 et s_2 est s_4 , et la restriction à S_c ;
 3. Si S ne satisfait pas à la condition 1., la condition 3. ne peut pas être satisfaite.

Ce fait ne devrait pas être étonnant : le quotient identifie les éléments de l'algèbre (Définition 5.1.4), de sorte qu'elle ne préserve pas le caractère libre d'une algèbre (Définition 5.1.2) et donc le caractère co-indépendant d'un ensemble (Définition 5.1.7). Pour le dire autrement, le quotient agit comme une restriction sur l'ensemble où s'applique l'ordre (§§A.3.3 et A.3.4, également la Définition 5.1.2), mais les propriétés de l'ordre correspondant à la co-indépendance ne sont pas préservées sous restriction.

Non seulement ce constat ouvre la voie à un certain nombre de questions techniques (telle : quelles propriétés et ensembles de propriétés sont préservées par quelles opérations ?), mais elle incite une réflexion philosophique.

Tout d'abord, il ne découle nullement de ce constat que les cadres définis dans le §5.1, et l'analyse de contrefactuels qui y est proposée, soit faux ou manque de plausibilité. *En pratique*, comme on l'a vu dans le paragraphe précédent, il y a des complications même avec les cadres_f et l'opération rigoureuse de fusion

entre elles. Si l'on vise à un modèle plus réaliste de la transition entre contextes, on pourrait permettre qu'il y a des transitions qui ne donnent pas lieu à des cadres au sens de la Définition 5.1.8, pour les moments où il n'est pas question de contrefactuels ou conséquences par exemple, et inversement qu'il y a des transitions dans des cadres, à partir des structures qui ne le sont pas. D'ailleurs, on pourrait même envisager d'employer la notion de fusion pour modéliser ces transitions, en employant des quotients supplémentaires pour assurer (ou éliminer) le caractère co-indépendant des ensembles appropriés.

En outre, il n'est pas clair qu'il soit *désirable* que les propriétés des cadres locaux (en l'occurrence les cadres au sens de la Définition 5.1.8) soient préservées par de telles opérations rigoureuses de transition. Les propriétés des cadres préservées par des opérations entrent sans doute dans une relation étroite avec les propriétés des opérations elles-mêmes, et notamment avec les règles logiques qu'elles valident. Mais les propriétés internes aux cadres relèvent de ce que l'on a appelé la « sémantique locale », les propriétés des transitions appartiennent à ce que l'on a appelé la logique des « étapes d'inférence » ; pourquoi doivent-elles être égales ? Étant donné que la perspective locale a permis la distinction entre ces deux domaines logiques, on ne saurait pas exiger leur identité sans examiner premièrement le sens et les conséquences d'une telle exigence. Il n'est pas le temps pour s'attarder sur cette question ; pour quelques autres remarques allant dans le même sens, voir la section suivante.

A.4 Règles logiques comme contraintes globales

La troisième manière de comprendre une « règle logique » lui rapproche d'une propriété structurelle *globale* du rapport entre le « langage » et le « monde ». À la différence de l'acceptation locale des règles logiques, selon laquelle elles expriment des aspects de la structure *interne* des cadres *locaux*, il s'agit ici des propriétés qui *transcendent* les cadres locaux particuliers, propriétés qui s'appliquent à tous les cadres, en tant qu'ils se rapportent les uns aux autres. On appelle une structure qui comporte toutes ces propriétés une « sémantique globale ». La question est de trouver, pour une collection de cadres locaux, une sémantique globale qui « résume » les cadres. Une « règle logique », dans cette acception, réfère à une propriété de cette sémantique globale.

Dans le premier paragraphe de cette section, on introduit la notion de sémantique globale et propose trois propriétés que l'on en souhaiterait. Dans le second paragraphe, on discute le problème de trouver une sémantique globale d'une collection de cadres locaux. Ce problème est complexe et difficile ;

on se contente de proposer une manière de la concevoir, comme la question de la construction d'un cadre qui entre dans des relations appropriées avec les cadres locaux de cette collection ; ce cadre ressemble aux cadres locaux employés jusqu'à présent, mais est, *grosso modo*, « plus grand » ou « gonflé ». Les considérations de la section précédente, et les outils techniques qui y sont développés, peuvent s'avérer utiles par rapport à cette question.

A.4.1 Qu'est qu'une sémantique globale

L'exemple, discuté dans (§5.3.1), de la suite de contrefactuels proposée par Lewis, sert à introduire la notion de sémantique globale.

$$A_1 \Rightarrow B \ ; \ A_1 \& A_2 \Rightarrow \neg B \ ; \ A_1 \& A_2 \& A_3 \Rightarrow B \ ; \ \dots$$

Il y a deux manières de comprendre cette suite : d'une part, avec une suite de cadres de plus en plus grand, dont chacun évalue seulement les premiers contrefactuels de la suite ; d'autre part, dans un cadre seul qui est dans un certain sens la « limite » de différents cadres appropriés aux différents contrefactuels, et qui assigne une valeur à tous les contrefactuels de la suite. Ce dernier « résumé » la suite de cadres locaux ; il fournit quelque chose de l'ordre d'une « sémantique globale » pour la suite de contrefactuels.

Ce que l'on appelle « sémantique globale » est donc de l'ordre d'une structure unique, qui, par rapport à une collection de cadres locaux (dont chacun fournit une sémantique locale pour son propre langage local) contient et évalue tous les énoncés apparaissant dans tous les cadres locaux, en respectant, autant que possible, les évaluations que reçoivent dans ces cadres locaux. Une des intentions de plusieurs théories du langage, voire de plusieurs « sémantiques », est apparemment de fournir une telle structure une, bien que, comme on l'a remarqué dans §5.3.2, il y ait une tension entre ce but et le fait que ces théories admettent des facteurs (telle une relation de similitude entre mondes) qui sont variables. Il n'est pas la place ici d'entrer dans une discussion détaillée des différentes théories du langage. Néanmoins, pour expliciter quelques propriétés souhaitables du rapport entre les cadres locaux, conçus comme les structures logiques à l'œuvre à des moments particuliers, et la sémantique globale, en tant que structure fixe qui « plane au-dessus » ces cadres locaux, tout en leur étant fidèle, les trois caractéristiques suivantes sont non seulement naturelles, mais renvoient à des aspects qui sont souvent exigés d'une sémantique.

Définition A.4.1. Soit $\mathcal{C}_i, i \in I$, une collection de cadres, avec comme langages locaux L_i et \mathcal{S}_G une « sémantique globale » pour $\bigcup_{i \in I} L_i$ – c'est-à-dire une

structure permettant l'évaluation de chaque énoncé qui appartient à L_i pour n'importe quel i . De plus, soit \sim une relation de congruence (appropriée) sur l'union disjoint de leurs éléments.

Concordance On dit que $\mathcal{S}_{\mathcal{G}}$ *concorde* avec $(\mathcal{C}_i)_{i \in I}$ si, pour tout énoncé A de $\bigcup_{i \in I} L_i$, l'évaluation accordée à A par $\mathcal{S}_{\mathcal{G}}$ est la même que l'évaluation accordée à A par les cadres \mathcal{C}_j qui contiennent A .

Respect de l'identité des énoncés On dit que $\mathcal{S}_{\mathcal{G}}$ *respecte l'identité des énoncés* si, pour toute paire de cadres \mathcal{C}_i et \mathcal{C}_j et toute paire d'éléments $A_i \in L_i$, $A_j \in L_j$, A_i et A_j sont identifiés dans $\mathcal{S}_{\mathcal{G}}$ si et seulement si $A_i \sim A_j$.

Minimalité On dit que $\mathcal{S}_{\mathcal{G}}$ est la sémantique globale *minimale* à avoir une certaine propriété relativement à $(\mathcal{C}_i)_{i \in I}$ si elle a cette propriété et, pour toute autre sémantique globale $\mathcal{S}_{\mathcal{G}}'$ avec cette propriété, tous les énoncés évalués par $\mathcal{S}_{\mathcal{G}}$ sont évalués par $\mathcal{S}_{\mathcal{G}}'$, et ils sont assignés la même évaluation.

De prime abord, il s'agit de trois propriétés de, ou si l'on préfère conditions sur, une sémantique globale : une sémantique globale pour une collection de cadres est une structure minimale ayant la propriété de concordance et de respect de l'identité des énoncés relative à cette collection. Quoiqu'elles répondent à des intuitions répandues, ces trois propriétés sont fort imprécises dans l'état actuel. En outre, il y a une certaine tension entre elles, dans la mesure où, pour toute collection de cadres où deux cadres évaluent un même énoncé (selon \sim) de manière différente, il n'y a pas de sémantique globale qui a la propriété de concordance et qui respecte l'identité des énoncés. Un des enjeux d'une étude sur la sémantique globale est de donner un sens plus rigoureux à ses intuitions, et, dans de tels cas de conflit, de garder autant que possible ces propriétés. Il s'agit, pour terminer cette annexe, d'ébaucher une manière de préciser ces contraintes, et plus généralement de concevoir la sémantique globale, en termes des cadres du type définie dans §5.1.2 (Définition 5.1.8). L'idée a été déjà évoquée dans la discussion de l'exemple de Lewis (§5.3.1) : on donne une sémantique globale pour une collection de cadres sous la forme d'un « cadre gonflé », qui ressemble à des cadres locaux à ceci près que, intuitivement, il est plus « grand », et qui entre dans des relations appropriées avec les cadres de la collection.

A.4.2 « Sémantique globale » comme cadre gonflé

La discussion précédente des propriétés désirables d'une sémantique globale rappelle la discussion des fusions des cadres dans §A.3.4. La propriété de concordance – selon laquelle la sémantique globale donne la même évaluation

aux énoncés, en l'occurrence aux conséquences, que les cadres locaux – rappelle la propriété de préservation des conséquences – selon laquelle la fusion de deux cadres accorde la même évaluation aux conséquences que les cadres originaux (Définition A.3.10). En outre, la tension entre la concordance et le respect de l'identité des énoncés se ressemble aux complexités qui se produisent dans des cas de fusion de deux cadres où la relation de congruence relie plusieurs paires d'énoncés entre eux. Il paraît donc recommandable d'employer les notions développées dans la section précédente pour préciser la question de la sémantique globale, et d'ailleurs pour en construire une. L'idée est simple : là où, dans la discussion des opérations sur les cadres qui reproduisent les règles logiques, il fallait des opérations qui « fusionnent » ou « combinent » deux cadres et qui satisfont à certaines propriétés, il faut, pour constituer une sémantique globale à partir d'une collection quelconque de cadres, il faut utiliser de telles opérations, ou des opérations apparentées, pour les « fusionnent » tous, en satisfaisant également à ces propriétés.⁵⁵

Concordance et préservation des conséquences

Pour exposer l'idée générale, il convient de laisser de côté pour un moment la propriété du respect de l'identité des énoncés, et de considérer la question (plus facile) de définir plus précisément la notion de sémantique globale comme structure minimale qui satisfait la propriété de concordance, et d'en donner une. On considère, comme ailleurs dans cette annexe, seulement les conséquences.

Dans §A.3.4, on a pu définir une relation \preceq entre les cadres (et entre les cadres_f) qui est telle que, pour le dire grossièrement, si les deux cadres entraînent dans cette relation, alors ils s'accorderaient sur les évaluations des conséquences qu'ils avaient « en commun » (Définition A.3.12 et Proposition A.3.3). Cette relation s'est avérée utile dans les considérations sur la préservation des conséquences par certaines opérations : puisque les cadres appropriés entraînent dans cette relation, les conséquences étaient préservées. Dans le même sens, on peut utiliser ce genre de relation pour définir une propriété d'un *cadre*, relative à une *collection de cadres*.

Définition A.4.2 (Concordance des cadres). Soit \mathcal{C}_i , $i \in I$, une collection de

⁵⁵On pourrait considérer que la différence entre le cas précédent et le cas de sémantique globale discutée ici est, du moins techniquement, celle entre le cas fini et le cas infini. Quoi qu'il en soit, on fera fi, pour l'essentiel, des questions techniques relatives à ce passage, car la discussion aura pour but seulement d'*introduire* la sémantique globale et de *motiver* une certaine manière de la concevoir et de l'étudier.

cadres. Le cadre \mathcal{C} *concorde avec* $(\mathcal{C}_i)_{i \in I}$ si,

$$\forall i \in I, \mathcal{C}_i \preceq \mathcal{C}$$

Il s'ensuit immédiatement de la Proposition A.3.3 que, si le *cadre* \mathcal{C} satisfait à la propriété de concordance des cadres (Définition A.4.2), alors il a la propriété de concordance au sens de la Définition A.4.1. Ce cadre, qui est intuitivement « grand » ou « gonflé », dans la mesure où il contient tous les énoncés figurant dans au moins un des cadres locaux, *caractérise* ou *résume* les conséquences satisfaites dans les cadres plus « petits » et « locaux », dans la mesure où il accorde à ces énoncés les mêmes valeurs de vérité qu'ils ont dans les cadres locaux. Ainsi, un tel *cadre gonflé* est une (version préliminaire de) « sémantique globale » pour la collection de cadres locaux.

Le diagramme suivant, où les cadres sont nommés \mathcal{C}_i , et une flèche de \mathcal{C}_i à \mathcal{C}_j représente le fait que $\mathcal{C}_i \preceq \mathcal{C}_j$, représente la concordance de cadres (Définition A.4.2) :

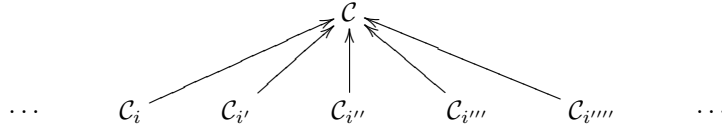


FIG. A.1 – Concordance des cadres

La propriété de concordance de cadres est donc une reformulation rigoureuse possible de la propriété de concordance (Définition A.4.1) ; également, un cadre « gonflé » ayant une telle propriété est un candidat naturel pour une sémantique globale. Il n'est pas difficile d'étendre cette conception aux autres propriétés apparentes d'une sémantique globale.

En effet, étant donné la relation \preceq , ou la notion d'application appropriée entre cadres, on pourrait définir une notion de cadre qui est minimal à posséder la propriété de concordance des cadres.

Définition A.4.3 (Minimalité pour la concordance de cadres). Un cadre \mathcal{C} est un cadre *minimal* ayant la propriété de concordance de cadres relative à une collection $(\mathcal{C}_i)_{i \in I}$ si, pour tout cadre \mathcal{C}' qui y satisfait pour la même collection, $\mathcal{C} \preceq \mathcal{C}'$.

Cette propriété reproduit l'intuition de la propriété intuitive de minimalité discutée précédemment : elle dit que le cadre minimal se plonge dans tout autre cadre ayant la propriété de concordance, si bien que, selon la Proposition A.3.3, cet autre cadre évalue tous les énoncés du cadre minimal, et de la même façon.

Les propriétés de minimalité et de concordance de cadres sont des types qui sont bien connus en mathématiques. En empruntant une terminologie à la théorie de catégories⁵⁶, qui vise à étudier de manière abstraite des systèmes avec ces genres de propriétés, on pourrait dire que la propriété de concordance de cadres (Définition A.4.2) exige que le cadre soit un *cône*, que la propriété supplémentaire de minimalité (Définition A.4.1) exige qu'elle est de plus *universel* à avoir cette propriété. Autrement dit, la première propriété consiste dans le fait pour le cadre d'entretenir les relations montrées dans le diagramme A.1 ; la seconde propriété consiste en le fait que, pour n'importe quel autre cadre qui entretient ces mêmes relations, il y a une flèche allant du cadre original à cet autre cadre (sur le diagramme A.1. En suivant toujours la terminologie de la théorie de catégories, on appelle un cadre ayant ces deux propriétés la *limite* de la collection $(\mathcal{C})_{i \in I}$ ⁵⁷.

En effet, il y a un nom pour le genre d'élément qui satisfait ces types de propriétés, à savoir « produit ». En l'occurrence, il s'agit du cadre qui est le produit (libre des cadres \mathcal{C}_i , fonction sur les cadres que l'on a déjà rencontré dans les considérations antérieures⁵⁸.

Fait A.4.1. *Pour $((\mathbf{B}, E)_i)_{i \in I}$ une collection de cadres, la limite de $((\mathbf{B}, E)_i)_{i \in I}$ est le produit*

$$(\mathbf{B}, E) = (\bigotimes_{i \in I} \mathbf{B}_i, \bigcup_{i \in I} E_i)$$

Autrement dit, le produit est le cadre minimal ayant la propriété de la concordance des cadres.

Ainsi, on arrive à proposer une sémantique globale (préliminaire), ayant les propriétés de concordance et de minimalité (ayant la propriété de concordance), en s'appuyant sur des opérations déjà introduites pour modéliser les règles logiques conçues comme propriétés des transitions entre cadres.

Respect de l'identité et la fusion

La propriété de concordance avec les cadres locaux ne suffit cependant pas pour faire une sémantique globale, parce que l'on souhaiterait d'une structure qui résume tous les cadres locaux qu'elle rend compte non seulement des toutes les évaluations des énoncés différents (en l'occurrence, les conséquences)

⁵⁶Voir par exemple MacLane (1998).

⁵⁷Dans la théorie de catégories, il y a une distinction entre la limite et la co-limite, dont on peut faire l'économie ici, dans la mesure où les deux notions sont duales.

⁵⁸Le produit libre a été explicitement défini sur les algèbres interprétées (Définition A.3.3) et sur les cadres_f (Définition A.3.7). Le fait que ce produit est bien défini sur les cadres, et son opération sur les ensembles co-indépendants, a été mentionné dans §A.3.5.

dans des différents cadres, mais également de l'identité des énoncés entre les cadres différents (représentée par la relation de congruence \sim ⁵⁹). Cette situation fait écho à celle qui motivait, dans §§A.3.3 et A.3.4, l'opération de *fusion* des cadres_f : il ne suffisait pas de prendre le produit (libre) de deux cadres, car il fallait de plus que la « fusion » tient compte de l'identité des énoncés entre les deux cadres (en tant que représentée par \sim). L'appui sur la discussion antérieure des cadres et des opérations entre eux, et l'idée de préciser la notion de sémantique globale en termes des relations et des applications entre cadres, et d'ailleurs d'en construire une à l'aide des opérations sur les cadres définis antérieurement, semble donc s'étendre au cas plus complexe où l'on demande qu'une sémantique globale satisfasse à toutes les trois propriétés de la Définition A.4.1.

Tout d'abord, on saurait proposer une formulation rigoureuse de la propriété du respect de l'identité des énoncés en termes des applications entre cadres.

Définition A.4.4 (Respecter l'identité des énoncés entre cadres). Soit une collection de cadres $(\mathcal{C}_i)_{i \in I}$ munie d'une relation de congruence appropriée, \sim , sur l'union disjoint de leurs éléments, et \mathcal{C} un cadre tel qu'il existe une application de chaque cadre \mathcal{C}_i dans \mathcal{C} . \mathcal{C} respecte l'identité des énoncés si, pour toute paire de cadres \mathcal{C}_i et \mathcal{C}_j et toute paire d'éléments $A_i \in \mathcal{C}_i$, $A_j \in \mathcal{C}_j$, les images de A_i et A_j dans \mathcal{C} coïncident si et seulement si $A_i \sim A_j$.

Le diagramme approprié pour un cadre qui satisfait les propriétés de la concordance et du respect de l'identité des énoncés est le suivant :

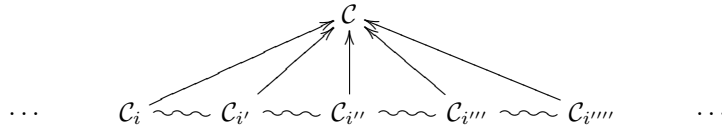


FIG. A.2 – Respect de l'identité des énoncés

De même que dans le cas de concordance considéré ci-dessus, un cadre minimal ayant ces deux propriétés est une limite de ce diagramme⁶⁰.

Or, comme on l'a mentionné auparavant, ce cas est beaucoup plus complexe et difficile que le cas où l'on demande seulement la concordance. Il n'est

⁵⁹On ne s'attarde pas ici sur la difficulté de préciser cette relation, difficulté que l'on a déjà discuté dans le contexte quelque peu différent de la théorie de la conscience ; voir §2.4 par exemple.

⁶⁰On omet la définition explicite de la minimalité par rapport aux deux propriétés, qui est une variation sur la Définition A.4.3.

pas toujours possible de trouver un cadre qui « fusionne » deux cadres, tout en respectant l'identité de leurs énoncés *et* en préservant leurs conséquences (§A.3.4, notamment l'exemple A.3.1); il n'est pas non plus possible d'avoir une sémantique globale qui satisfasse *simultanément* la propriété du respect de l'identité des énoncés et la propriété de la concordance (§A.4.1). Pour le dire dans les termes introduits ci-dessus, il n'y a pas toujours de limite du diagramme A.2 lorsque les flèches sont interprétées comme des relations \preceq , ou, autrement dit, comme des applications injectives entre cadres (Définition A.3.12).

Tout ceci n'a pas pour conséquence l'impossibilité de sémantique globale, mais plutôt la nécessité, pour préciser la notion de sémantique globale, et d'ailleurs pour en trouver une, de compromettre entre les propriétés de concordance et de respect de l'identité des énoncés. Ceci ne justifie pas non plus quelconque rejet de la conception de la question qui vient d'être proposée. Au contraire, l'idée d'une sémantique globale fournie par un cadre qui entre dans des rapports appropriés avec les cadres pertinents, et qui peut d'ailleurs être construit à partir de ces cadres avec des opérations appropriées, reste féconde.

En premier lieu, cette idée permet une formulation claire du problème de la sémantique globale. Le diagramme A.2, a-t-on remarqué, représente les propriétés souhaitées d'une sémantique globale : elle est le cadre minimal qui entre dans la bonne relation avec les cadres \mathcal{C}_i (représentées par les flèches), tout en respectant l'identité des énoncés (représentée par les lignes onduleuses entre les \mathcal{C}_i). L'incompatibilité de la concordance et du respect de l'identité des énoncés, entendues en sens strict, implique que la relation pertinente (les flèches) ne peut pas être \preceq (ou l'application injective). La question de la sémantique globale pourrait donc être posée de façon suivante : quelle relation entre \mathcal{C} et les \mathcal{C}_i permettrait de considérer \mathcal{C} , la limite de ce diagramme, comme une sémantique globale ? Autrement dit : comment interpréter les flèches dans ce diagramme pour que la limite soit une sémantique globale ? Les relations cherchées doivent conserver autant que possible l'évaluation des conséquences dans des cadres, respecter autant que possible l'identité des énoncés (représentée par \sim), mais permettre que des limites, c'est-à-dire des sémantiques globales, existent.

Alors que la considération approfondie de cette question déborde les limites du présent travail, la discussion précédente permet d'ébaucher quelques possibilités. Plus précisément, tout en soulignant qu'il pourrait y avoir plusieurs possibilités de relations entre sémantiques globales et cadres, et donc plusieurs *notions* de sémantique globale, l'opération de fusion introduite dans §A.3.4 fournit un candidat naturel et recommandable pour une sémantique globale. L'intuition soutenant cette opération est simple : le produit libre *met ensemble* les cadres

sans identifier leurs énoncés entre eux, ensuite le quotient *identifie* les énoncés appropriés (§A.3.4). D'ailleurs, cette intuition paraît également pertinente pour les sémantiques globales. Le cadre obtenu en *ajoutant* tous les cadres locaux les uns aux autres, sans faire attention aux identifications entre les énoncés, pour ensuite *identifier* tous les énoncés appropriés, a du moins une prétention à être une sémantique globale.

Techniquement, pour une collection de cadres_f (et donc *a fortiori* de cadres⁶¹) on pourrait proposer comme sémantique globale la fusion des cadres de cette collection, dont la définition, inspirée de la Définition A.3.9 est comme suit.

Définition A.4.5. Soit une collection de cadres_f $((\mathbf{B}, \prec)_i)_{i \in I}$, avec une relation de équivalence \sim entre leurs énoncés. La *fusion*, $(\mathbf{B}, \prec)_I$, a

$$\mathbf{B}_I = \bigotimes_{i \in I} \mathbf{B}_i / (\bigvee_{A^i \sim A^j} (A^i \triangle A^j))$$

$$\prec = \prod_{i \in I} \prec_i \mid \prod_{i \in I} S_i \setminus \{s \in \prod_{i \in I} \mid s \leq \bigvee_{A^i \sim A^j} (A^i \triangle A^j)\}$$

Cette proposition de sémantique globale approche du problème de l'autre côté. La fusion pourrait être conçue comme la limite d'un diagramme du même type que le diagramme A.2; pour le concevoir ainsi, il faut préciser une interprétation des flèches du diagramme, ou autrement dit les relations appropriées entre la fusion et les cadres locaux particuliers. Une considération approfondie de cette question développerait sans doute la Remarque A.3.6 de §A.3.4 qui mettait en évidence les effets de l'opération de quotient sur l'évaluation des conséquences. En effet, il s'agirait de développer des notions de préservation des conséquences et de concordance, qui sont plus subtiles que celles des Définitions A.3.10 et A.4.1, et qui sont telles que la fusion est le cadre minimal ayant ces propriétés. Cette considération doit cependant être reportée à un travail ultérieur.

Et les règles logiques? Si l'on admet la proposition de sémantique globale comme cadre « gonflé » – ressemblant aux cadres « ordinaires » mais plus « grand » –, lequel est un cadre minimal entrant dans des rapports appropriés avec les cadres locaux, et lequel pourrait éventuellement être construit à partir de ces derniers à l'aide des opérations définies, alors on peut tirer des conclusions relatives aux règles logiques comme de simples *corollaires* des considérations des deux sections précédentes.

D'une part, les propriétés internes des cadres, et des cadres_f, telles que les règles logiques qui s'y appliquent, ont été déjà étudiées (§§A.1, A.2 et A.3.1). Donc, le fait d'utiliser un cadre (ou un cadre_f) comme sémantique globale a

⁶¹On en viendra prochainement à l'importance de l'usage des cadres_f plutôt que des cadres.

pour conséquence que l'on connaît immédiatement certaines des règles logiques qui s'y appliquent ; notamment, les règles **(Réf)-(Or)**, entendues dans le sens approprié, s'appliquent aux deux types de structure (§A.2 et Remarque A.3.3).

D'autre part, si l'on construit la sémantique globale par l'application des opérations, les considérations de §A.3, relatives aux propriétés des algèbres, et des ordres – bref, des cadres – qui sont préservées sous différentes opérations, deviennent pertinentes. Et les propriétés qui sont préservées ont des conséquences sur les règles qui, valides à l'intérieur des cadres locaux, sont toujours valides dans le cadre « gonflé » qui résume tous ces cadres locaux. La nécessité d'employer le cadre_f dans la considération de l'opération de fusion, et donc éventuellement dans la construction d'une sémantique globale à l'aide de cette opération, relève du fait que certaines propriétés des cadres ne sont pas préservées par cette opération (§A.3.5). En outre, les remarques faites dans §A.3.5, à l'égard de la relation entre les propriétés internes aux cadres et les transitions entre eux, sont également pertinentes pour la discussion de la sémantique globale. De même qu'il est une question subtile de savoir si les propriétés logiques « actives » à des moments particuliers (les propriétés des cadres locaux) sont, ou devraient être, celles qui sont préservées dans les transitions d'un moment à l'autre (préservées par les opérations sur les cadres), de même il n'est pas claire si l'on devrait demander qu'une sémantique globale ait les mêmes propriétés que les cadres locaux qu'elle résume. Autrement dit, il reste à savoir dans quelle mesure les règles logiques qui s'appliquent à un niveau local, sont, ou devraient être les mêmes que celles qui « transcendent » ce niveau local, qui s'appliquent à un niveau plus global.

Pour dire le moins, *il ne s'ensuit aucunement* de la décision de ne pas s'occuper des règles logiques dans les Chapitres 4 et 5, que le cadre théorique qui y est proposé est *incapable* de rendre compte des règles logiques. Au contraire, on a montré dans cette annexe qu'il offre non seulement trois interprétations *rigoureuses* des règles logiques, qui entraînent des questions philosophiques et techniques nouvelles et intéressantes, mais également des outils convenables à l'étude détaillée de ces questions.

Bibliographie

- Adams, E. W. (1970). Subjunctive and indicative conditionals. *Foundations of Language*, 6(1) :89–94.
- Adams, E. W. (1975). *The Logic of Conditionals*. D. Reidel Publishing Co., Dordrecht.
- Anderson, A. R. and Belnap, N. D., editors (1975). *Entailment : The Logic of Relevance and Necessity*, volume I. Princeton University Press, Princeton.
- Anderson, J. R. (1983). *The Architecture of Cognition*. Harvard UP, Cambridge, MA.
- Aristote (1953). *Metaphysics*. Clarendon Press, Oxford. Trad par W. D. Ross.
- Aristote (1980). *La Poétique*. Ed du Seuil. Trad par R. Dupont-Roc & J. Lallot.
- Aristote (1990). *Poétique*. Librairie Générale Française. Trad par M. Magnien.
- Aristote (1991). *Métaphysique*. Vrin. Trad par J. Tricot.
- Austin, J. L. (1979). *Philosophical Papers*. OUP, Oxford. Trad fr, *Ecrits Philosophiques*, Éd de Seuil, 1994.
- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. CUP, Cambridge.
- Baddeley, A. D. (1992). Is working memory working? the fifteenth bartlett lecture. *Quarterly Journal of Experimental Psychology*, 44A :1–31.
- Baker, G. P. and Hacker, P. M. S. (1983a). *An Analytic Commentary on the Philosophical Investigations*. Blackwell, Oxford.
- Baker, G. P. and Hacker, P. M. S. (1983b). *Wittgenstein : Meaning and Understanding*. Blackwell, Oxford.
- Baker, G. P. and Hacker, P. M. S. (1985). *Wittgenstein : Rules, Grammar and Necessity*. Blackwell, Oxford.
- Baltag, A. and Moss, L. S. (2004). Logic for epistemic programs. *Synthese*, 60 :1–59.
- Baltag, A., Moss, L. S., and Solecki, S. (1998). The logic of common knowledge, public announcements and private suspicions. In Gilboa, I., editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, pages 43–56.

- Banfield, A. (1982). *Unspeakable Sentences : narration and representation in the language of fiction*. Routledge & Kegan Paul, Boston.
- Barthes, R. (1973). *Le plaisir du texte*. Éd du Seuil.
- Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT Press, Cambridge, MA.
- Bedford, E. (1957). Emotions. *Proceedings of the Aristotelian Society*, pages 281–304. Aussi dans Gustafson D. (éd) *Essays in Philosophical Psychology*, New York : Doubleday, 1967, dont est pris la pagination.
- Benacerraf, P. (1983). What numbers could not be. In Benacerraf and Putnam (1983).
- Benacerraf, P. and Putnam, H. (1983). *Philosophy of Mathematics ; Selected Readings*. CUP, Cambridge.
- Bennett, J. F. (1974). Counterfactuals and possible worlds. *Canadian Journal of Philosophy*, 4 :381–402.
- Bergson, H. (1948). *Essai sur les Données Immédiates de la Conscience*. PUF.
- Bergson, H. (1959). *Matière et Mémoire*. PUF.
- Blackburn, P., Rijke, M. d., and Venema, Y. (2001). *Modal Logic*. Cambridge University Press, Cambridge.
- Block, N., editor (1980). *Readings in the Philosophy of Psychology*, volume 1. Methuen, London.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18 :227–247.
- Block, N. (1996). Functionalism. In *The Encyclopedia of Philosophy Supplement*. Macmillan, London.
- Block, N. (1999). Holism, mental and semantic. In *The Encyclopedia of Philosophy*. Routledge, London.
- Block, N. and Fodor, J. A. (1972). What psychological states are not. *Philosophical Review*, 81 :159–181. Aussi dans Block (1980) dont est pris la pagination.
- Blond, J. M. L. (1970). *Logique et méthode chez Aristote ; étude sur la recherche des principes dans la physique aristotélicienne*. J. Vrin.
- Boolos, G. (1971). The iterative conception of set. *The Journal of Philosophy*, 68 :215–232.
- Bouveresse, J. (1976). *Le Mythe de l'Intériorité*. Éd de Minuit.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4 :73–121.
- Caillouis, R. (1967). *Les jeux et les hommes*. Gallimard. (1^{re}éd 1958).

- Carnap, R. (1928). *Der logische Aufbau der Welt*. Mainer, Hamburg. trad fr *La construction logique du monde*, 2002, Vrin.
- Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago.
- Carnap, R. (1956). Empiricism, semantics and ontology. In *Meaning and Necessity*. University of Chicago Press, Chicago. aussi dans Benacerraf & Putnam, 1983.
- Carroll, L. (1895). What the tortoise said to achilles. *Mind*, 4 :278–280.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press, Oxford.
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, 55 :289–307.
- Cohn, D. (1999). *The Distinction of Fiction*. John Hopkins University Press, Baltimore. Trad fr *Le Propre de la Fiction*, Éd du Seuil, 2001.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104 :163–191.
- Cowan, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition*, 21 :162–167.
- Currie, G. (1988). Fictional names. *Australian Journal of Philosophy*, 66 :471–90.
- Currie, G. (1990). *The Nature of Fiction*. CUP, Cambridge.
- Currie, G. (1995). Imagination and simulation : Aesthetics meets cognitive science. In Davies and Stone (1995).
- Davidson, D. (1976). Hume's cognitive theory of pride. In Davidson (1980). Trad fr *Actions et événements* 1993, PUF.
- Davidson, D. (1980). *Essays on Actions and Events*. OUP, Oxford. Trad fr *Actions et événements* 1993, PUF.
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.
- Davies, M. and Stone, T., editors (1995). *Mental Simulation*. Blackwell, Oxford.
- Dennett, D. C. (1977). Revue de fodor, 1975. *Mind*, 86 :265–280.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press, Cambridge, MA. Trad fr *La Stratégie de l'interprète*, 1990, Gallimard.
- Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown & Company.
- Dretske, F. I. (1970). Epistemic operators. *Journal of Philosophy*, 67 :1007–1023.

- Dunn, J. M. (1986). Relevance logic and entailment. In Gabbay, D. and Guenther, F., editors, *Handbook of Philosophical Logic*, volume III, pages 117–224. Dordrecht, Reidel.
- Eco, U. (1965). *L'œuvre ouverte*. Éd du Seuil.
- Eco, U. (1979). *The Role of the Reader*. Indiana University Press, Bloomington.
- Eco, U. (1985). *Lector in fabula*. Éd Grasset & Fasquelle.
- Edgington, D. (1995). On conditionals. *Mind*, 104 :235–329.
- Elster, J. (1999a). *Alchemies of the Mind : Rationality and the Emotions*. CUP, Cambridge.
- Elster, J. (1999b). *Strong Feelings*. MIT Press, Cambridge.
- Engel, P., editor (2000). *Believing and Accepting*. Kluwer, Dordrecht.
- Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1995). *Reasoning about Knowledge*. MIT Press, Cambridge, MA.
- Feagin, S. (1996). *Reading with Feeling*. Cornell University Press, Ithaca.
- Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30 :265–300.
- Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press, Cambridge MA.
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Brain and Behavioral Sciences*, 3 :63–110.
- Fodor, J. A. (1985). Fodor's guide to mental representation. *Mind*, XCIV :76–100.
- Frege, G. (1892). Über begriff und objekt. *Vierteljahrsschrift für wissenschaftliche Philosophie*, XVI :192–205.
- Frege, G. (1893). *Grundesetze der Arithmetik*. Geord Olms, Hildesheim.
- Frege, G. (1969). Logik. In *Nachgelassene Schriften*. F. Meiner, Hamburg. 1897. Trad fr *Ecrits Posthumes*, 1994, J. Chambon.
- Frege, G. and Hilbert, D. (1900). La correspondance frege/hilbert. In Rivenc and de Rouilhan (1992).
- Gabbay, D. (1972). A general theory of the conditional in terms of a ternary operator. *Theoria*, 38 :97–104.
- Gärdenfors, P. (1978). On the logic of relevance. *Synthese*, 37 :351–367.
- Gärdenfors, P. (1986). Belief revisions and the ramsey test for conditionals. *The Philosophical Review*, 95 :81–93.
- Gärdenfors, P. (2000). *Conceptual Spaces*. MIT Press, Cambridge, MA.
- Gennette, G. (1991). *Fiction et Diction*. Éd du Seuil.

- Goldie, P. (2000). *The Emotions : A Philosophical Exploration*. Clarendon Press, Oxford.
- Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, 4 :161–85.
- Goldman, A. (1992). In defense of simulation theory. *Mind and Language*, 7 :104–119.
- Goldman, A. (1995). Empathy, mind and morals. In Davies and Stone (1995).
- Goodman, N. (1954). *Fact, Fiction and Forecast*. Harvard University Press, Cambridge, MA. 4ème édn 1983, Cambridge : Harvard University Press. Trad fr, *Faits, Fictions et Prédications*, Editions de Minuit, 1984.
- Goodman, N. (1968). *Languages of Art*. Hackett, Indiana. Trad fr *Langages de l'art*, J. Chambon, 1990.
- Goodman, N. (1978). *Ways of Worldmaking*. Hackett, Indiana. trad fr *Manières de faire les mondes*, J. Chambon, 1992.
- Gordon, R. (1986). Folk psychology as simulation. *Mind and Language*, 1 :158–70.
- Gordon, R. (1987). *The Structure of Emotions : Investigations in Cognitive Philosophy*. CUP, Cambridge.
- Gordon, R. (1995). Simulation without introspection or inference from me to you. In Davies and Stone (1995).
- Graham, G. (2005). Behaviorism. In *The Stanford Encyclopedia of Philosophy*.
- Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press, Cambridge.
- Groendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, 14 :39–100.
- Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic*, 17 :157–170.
- Gärdenfors, P. (1988). *Knowledge in Flux : Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA.
- Gärdenfors, P. and Rott, H. (1995). Belief revision. In Gabbay, Hogger, and Robinson, editors, *Handbook of Logic in AI and Logic Programming*, volume 4. Oxford University Press, Oxford.
- Hacker, P. M. S. (1993). *Wittgenstein : Meaning and Mind*. Blackwell, Oxford.
- Hacking, I. (1979). What is logic? *Journal of Philosophy*, 76 :285–319.
- Hamburger, K. (1957). *Die Logik der Dichtung*. E. Klett, Stuttgart. Trad fr *Logique des genres littéraires*, Éd du Seuil, 1986.
- Hansson, S. O. (2003). Ten philosophical problems in belief revision. *Journal of Logic and Computation*, 13 :37–49.

- Harper, W. L. (1975). Rational belief change, popper functions and counterfactuals. *Synthese*, 30 :221–262.
- Harris, P. (1992). From simulation to folk psychology : the case for development. *Mind and Language*, 7 :120–144.
- Hausdorff, F. (1914). *Grundzüge der Mengenlehre*. Von Veit, Leipzig.
- Heal, J. (1986). Replication and functionalism. In Butterfield, J., editor, *Language, Mind and Logic*. CUP, Cambridge.
- Henkin, L., Monk, J. D., and Tarski, A. (1985a). *Cylindrical Algebras. Part I*. North-Holland.
- Henkin, L., Monk, J. D., and Tarski, A. (1985b). *Cylindrical Algebras. Part II*. North-Holland.
- Hilbert, D. (1899). *Grundlagen der Geometrie*. B. G. Teubner, Leipzig.
- Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4 :475–484.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. (1989). *Induction*. MIT Press, Cambridge.
- Huizinga, J. (1951). *Homo ludens : Essai sur la fonction sociale du jeu*. Gallimard.
- Husserl, E. (1913a). Ideen zu einer reinen phänomenologie und phänomenologischen philosophie. In *Jahrbuch für Philosophie und phänomenologische Forschung*, 1. Max Niemeyer, Halle. Trad fr *Idées directrices pour une phénoménologie*, 1950, Gallimard.
- Husserl, E. (1913b). *Logische Untersuchungen*. 2^e edition. Trad fr *Recherches Logiques*, 1992-1994, PUF.
- Husserl, E. (1928). Vorlesungen zur phänomenologie des inneren zeitbewusstseins. In Heidegger, M., editor, *Jahrbuch für Philosophie und phänomenologische Forschung*, 9. Max Niemeyer, Halle. Trad fr *Leçons pour une phénoménologie de la conscience intime du temps*, 1^{re} éd 1964, 2^e éd 2002, PUF.
- Ingarden, R. (1965). *Das literarische Kunstwerk. Eine Untersuchung aus dem Grenzgebiet des Ontologie, Logik und Literaturwissenschaft*. M. Niemeyer, Tübingen. 3^e éd. Trad fr *L'oeuvre d'art littéraire*, 1983, Lausanne : L'Age d'homme (pagination française).
- Ingarden, R. (1968). *Vom Erkennen des literarischen Kunstwerks*. M. Niemeyer, Tübingen. Trad angl, *The Cognition of the Literary Work of Art*, 1973, Evanston, Illinois : Northwestern University Press (pagination anglaise).
- Iser, W. (1976). *Der Akt des Lesens*. W. Fink, München. Trad angl, *The Act of Reading*, 1978, Baltimore, Maryland : John Hopkins University Press (pagination anglaise).

- Iser, W. (1991). *Das Fictif und das Imaginäre : Perspektiven literarischer Anthropologie*. Suhrkamp, Frankfurt a. M. Trad angl *The Fictive and the Imaginary : Charting Literary Anthropology*, 1993, Baltimore, Maryland : John Hopkins University Press (pagination anglaise).
- James, W. (1884). What is an emotion ? *Mind*, 9 :188–205.
- James, W. (1981). *Principles of Psychology*. Harvard University Press, Cambridge, MA.
- Johnson-Laird, P. N. (1983). *Mental Models : Towards a Cognitive Science of Language, Inference and Consciousness*. CUP, Cambridge.
- Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. CUP, Cambridge.
- Kamp, H. (1981). The paradox of the heap. In Mönnich, U., editor, *Aspects of Philosophical Logic*, pages 225–277. Reidel, Dordrecht.
- Kant, E. (1781). *Kritik der reinen Vernunft*. 2^eéd, 1787. Trad fr *Critique de la Raison Pure*, G-F Flammarion, 2001.
- Kenny, A. (1963). *Action, Emotion and Will*. Routledge, London.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension : A construction-integration model. *Psychological Review*, 95 :163–182.
- Konieczny, S. and Pérez, R. P. (2000). A framework for iterated revision. *Journal of Applied Non-Classical Logics*, 10(3-4) :339–367.
- Krasner, D. A. (2002). Semantics and fiction. *Erkenntnis*, 57 :259–275.
- Kratzer, A. (1981). Partition and revision : The semantics of counterfactuals. *Journal of Philosophical Logic*, 10 :201–216.
- Kraus, S., Lehmann, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44 :167–207.
- Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language*. Blackwell, Oxford.
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago UP, Chicago, 2^eédition. Trad fr *La Structure des Révolutions Scientifiques*, 1983, Gallimard (pagination française).
- Kuratowski, K. (1966). *Topology*, volume I. PWN - Polish Scientific Publishers, Academic Press, Warsaw, New York and London.
- Kvart, I. (1986). *A Theory of Counterfactuals*. Hackett, Indianapolis.
- Lehmann, D. and Magidor, M. (1992). What does a conditional knowledge base entail ? *Artificial Intelligence*, 55 :1–60.
- Leibniz, G. W. (1966). *Nouveaux Essais sur l'Entendement Humain*. Garnier-Flammarion.
- Leibniz, G. W. (1969). *Opusculs Philosophiques Choisis*. J. Vrin.

- Leibniz, G. W. (1998). *Recherches Générales sur l'Analyse des Notions et des Vérités*. PUF.
- Leslie, A. (1987). Pretense and representation : the origins of 'theory of mind'. *Psychological Review*, 94 :412–26.
- Lewis, D. K. (1970). General semantics. *Synthese*, 22 :18–67.
- Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50 :249–258.
- Lewis, D. K. (1973a). Causation. *Journal of Philosophy*, 70 :556–567.
- Lewis, D. K. (1973b). *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts.
- Lewis, D. K. (1973c). Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2(4) :418–446.
- Lewis, D. K. (1974). Radical interpretation. *Synthese*, 23 :331–344. Lewis 1983 (dont est pris la pagination).
- Lewis, D. K. (1975). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85 :297–315.
- Lewis, D. K. (1978). Truth in fiction. *American Philosophical Quarterly*, 15 :37–46. aussi dans Lewis 1983, dont est pris la pagination.
- Lewis, D. K. (1979a). Attitudes de dicto and de se. *The Philosophical Review*, 88 :513–543.
- Lewis, D. K. (1979b). Counterfactual dependence and time's arrow. *Noûs*, 13 :455–476.
- Lewis, D. K. (1979c). Scorekeeping in the language game. *Journal of Philosophical Logic*, 8 :339–359.
- Lewis, D. K. (1981). Ordering semantics and premise semantics for conditionals. *Journal of Philosophical Logic*, 10(2) :217–234. aussi dans Lewis 1998.
- Lewis, D. K. (1983). *Philosophical Papers*, volume 1. OUP, Oxford.
- Lewis, D. K. (1986a). *On the Plurality of Worlds*. Blackwell, Oxford.
- Lewis, D. K. (1986b). *Philosophical Papers*, volume 2. OUP, Oxford.
- Lewis, D. K. (1986c). A subjectivist's guide to objective chance. In Lewis (1986b), pages 83–132.
- Lewis, D. K. (1991). *Parts of Classes*. Blackwell, Oxford.
- Lewis, D. K. (1994). Humean supervenience debugged. *Mind*, 104 :473–490.
- Lewis, D. K. (1998). *Papers in Philosophical Logic*. CUP, Cambridge.
- Loar, B. (1981). *Mind and Meaning*. CUP, Cambridge.
- Lowe, E. J. (1995). The truth about counterfactuals. *The Philosophical Quarterly*, 45 :41–59.

- MacLane, S. (1998). *Categories for the Working Mathematician*. Springer-Verlag, New York, 2^e edition.
- Makinson, D. (2005). *Bridges from Classical to Nonmonotonic Logic*. King's College Publications, London.
- Mares, E. D. (2004). *Relevant Logic : A Philosophical Interpretation*. CUP, Cambridge.
- Mares, E. D. and Fuhrmann, A. (1995). A relevant theory of conditionals. *Journal of Philosophical Logic*, 24 :645–665.
- McKinsey, J. C. C. and Tarski, A. (1944). The algebra of topology. *Annals of Mathematics*, 45 :141–191.
- Meinong, A. (1904). Über gegenstandstheorie. *Gesamtausgabe*, II, 1968–1978. Trad fr dans *Théorie de l'Objet et Présentation Personnelle*, J. Vrin, 1999 (pagination allemande).
- Monk, J. D. (2000). An introduction to cylindric set algebras. *Logic Journal of IGPL*, 8 :451–492.
- Nayak, A. C. (1994). Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41 :353–390.
- Nichols, S. and Stich, S. P. (2003). *Mindreading*. Clarendon, Oxford.
- Nozick, R. (1981). *Philosophical Explanations*. Belnap Press, Cambridge, MA.
- Parsons, T. (1975). A meinongian analysis of fictional objects. *Grazer Philosophische Studien*, 1 :73–86.
- Pavel, T. (1986). *Fictional Worlds*. Harvard University Press, Cambridge, MA. Trad fr *Univers de la fiction*, Éd du Seuil, 1988 (pagination française).
- Pavel, T. (2003). *La pensée du roman*. Gallimard.
- Plagnol, A. (1993). *Élaboration d'un modèle de désorganisation des représentations mentales par décontextualisation fonctionnelle de l'information*. thèse de doctorat de sciences cognitives, Université de Paris 11, Orsay.
- Plagnol, A. (2004). *Espaces de représentation*. CNRS Editions.
- Platon (2002). *La République*. Flammarion.
- Popper, K. R. (1990). *A World of Propensities : Two New Views of Causality*. Thoemmes, Bristol. Trad fr *Un univers de propensions*, L'éclat, 1992.
- Potts, G. R. and Peterson, S. B. (1985). Incorporation versus compartmentalization in memory for discourse. *Journal of Memory and Language*, 24 :107–118.
- Pouivet, R. (2006). *Le réalisme esthétique*. PUF.
- Putnam, H. (1975). The meaning of 'meaning'. In *Mind, Language and Reality*. CUP, Cambridge.
- Quine, W. V. O. (1940). *Mathematical Logic*. Harvard University Press, Cambridge, MA.

- Quine, W. V. O. (1951). Two dogmas of empiricism. In Quine (1953).
- Quine, W. V. O. (1953). *From a Logical Point of View*. Harvard University Press, Cambridge, MA.
- Quine, W. V. O. (1960). *Word and Object*. MIT Press, Cambridge, MA.
- Quine, W. V. O. (1964). Ontological reduction and the world of numbers. *Journal of Philosophy*, 61 :209–216.
- Quine, W. V. O. (1969). *Ontological Relativity, and Other Essays*. Columbia University Press, New York.
- Quine, W. V. O. (1981). *Theories and Things*. Belnap Press, Cambridge, MA.
- Quine, W. V. O. (1992). *Pursuit of Truth*. Harvard University Press.
- Ramsey, F. P. (2003). *Logique Philosophique et Probabilités*. J. Vrin.
- Recanati, F. (2000). *Oratio Obliqua, Oratio Recta : An Essay on Metarepresentation*. MIT Press, Cambridge, MA.
- Ricœur, P. (1983). *Temps, Récit et Fiction*, volume I. Ed du Seuil.
- Rivenc, F. and de Rouilhan, P., editors (1992). *Logique et Fondements des Mathématiques ; Anthologie (1890-1914)*. Payot.
- Rott, H. (2003). Coherence and conservatism in the dynamics of belief. part ii : Iterated belief change without dispositional coherence. *Journal of Logic and Computation*, 13 :111–145.
- Rott, H. (2004). A counterexample to six fundamental principles of belief formation. *Synthese*, 139 :225–240.
- Russell, B. (1937). *Principles of Mathematics*. Allen & Unwin, London, 2^e edition.
- Russell, B. (1962). *An Inquiry into Meaning and Truth*. Penguin, London.
- Rybakov, V. V. (1997). *Admissibility of Logical Inference Rules*. Elsevier Science B. V., Amsterdam.
- Ryle, G. (2000). *The Concept of Mind*. Penguin, London. 1^{er} éd. 1949. Trad. fr. *La notion d'esprit*, Payot, 1978 (pagination anglaise).
- Sambin, G. (2003). Some points in formal topology. *Theoretical Computer Science*, 305 :347–408.
- Sartre, J.-P. (1943). *L'être et le néant*. Gallimard.
- Sartre, J.-P. (1986). *L'imaginaire*. Gallimard. 1^{er} éd. 1940.
- Sartre, J.-P. (1995). *Esquisse d'une théorie des émotions*. Hermann. 1^{er} éd. 1938.
- Savage, L. (1954). *The Foundations of Statistics*. Dover, New York. 2^e éd. 1971.

- Schacter, D. L. (1989). On the relation between memory and consciousness : Dissociable interactions and conscious experience. In Roediger, H. L. and Craik, F. I. M., editors, *Varieties of memory and consciousness : Essays in honour of Endel Tulving*. LEA, Hillsdale, NJ.
- Schaeffer, J.-M. (1987). Fiction, feint et narration. *Critique*, 481-482 :555–76.
- Schaeffer, J.-M. (1999). *Pourquoi la Fiction ?* Ed du Seuil.
- Searle, J. R. (1969). *Speech Acts : An Essay in the Philosophy of Language*. CUP, Cambridge. Trad fr *Les actes de langage*, Hermann, 1972.
- Searle, J. R. (1975). The logical status of fictional discourse. *New Literary History*, VI. Aussi dans Searle 1979.
- Searle, J. R. (1979). *Expression and Meaning : studies in the theory of speech acts*. CUP, Cambridge. Trad fr *Sens et Expression*, Éd de Minuit, 1982.
- Segeberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39 :287–306.
- Sellars, W. (1997). *Empiricism and the Philosophy of Mind*. Harvard UP, Cambridge.
- Shapiro, S. (1998). Logical consequence : Models and modality. In Schirn, M., editor, *Philosophy of mathematics today : Proceedings of an international conference in Munich*, pages 131–156, Oxford.
- Sher, G. (1991). *The Bounds of Logic*. MIT Press, Cambridge, MA.
- Simons, P. (1987). *Parts. A Study in Ontology*. Clarendon Press, Oxford.
- Solomon, R. C. (1973). Emotion and choice. *The Review of Metaphysics*, XVII. aussi dans *Explaining Emotions*, Rorty A. O. (éd), Berkeley : University of California Press, 1980, dont est pris la pagination.
- Stalnaker, R. C. (1968). A theory of conditionals. In Rescher, N., editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell Publishers, Oxford. aussi dans *Causation and Conditionals* 1975, E Sosa (éd), Oxford : OUP (dont est pris la pagination).
- Stalnaker, R. C. (1975). Indicative conditionals. *Philosophia*, 5 :269–286.
- Stalnaker, R. C. (1980). A defense of conditional excluded middle. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs : Conditionals, Belief, Decision, Chance, and Time*, pages 87–104. D. Reidel Publishing Co., Dordrecht.
- Stalnaker, R. C. (1984). *Inquiry*. MIT Press, Cambridge, MA.
- Stalnaker, R. C. (1986). Possible worlds and situations. *Journal of Philosophical Logic*, 15 :109–123.
- Stalnaker, R. C. (1999). *Context and Content : Essays on Intentionality in Speech and Thought*. Oxford University Press, Oxford.
- Stempel, W.-D. (1979). Aspects génériques de la réception. *Poétique*, 39.

- Stich, S. and Nichols, S. (1995). Folk psychology : Simulation or tacit theory. In Davies and Stone (1995).
- Sutherland, W. A. (1975). *Introduction to Metric and Topological Spaces*. OUP, Oxford.
- Tarski, A. (1944). The semantic conception of truth and the foundation of semantics. *Philosophy and Phenomenological Research*, 4.
- Tarski, A. (1956a). *Logic, Semantics and Metamathematics*. Clarendon Press, Oxford.
- Tarski, A. (1956b). On the concept of logical consequence. In Tarski (1956a).
- Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, 7 :143–54.
- Todorov, T. (1970). *Introduction à la littérature fantastique*. Éd du Seuil.
- Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211 :453–458.
- Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59 :251–278.
- Tye, M. (1994). Sorites paradoxes and the semantics of vagueness. In Tomberlin, J., editor, *Philosophical Perspectives : Logic and Language*. Ridgeview, Atascadero, California.
- van Benthem, J. (1996). *Exploring Logical Dynamics*. CSLI Publications, Stanford.
- van Benthem, J. and ter Meulen, A., editors (1997). *Handbook of Logic and Language*. Elsevier Science B. V., Amsterdam.
- van Frassen, B. (1966). Singular terms, truth-value gaps, and free logic. *Journal of Philosophy*, 63 :481–495.
- Walton, K. L. (1990). *Mimesis as Make-Believe : On the Foundations of the Representative Arts*. Harvard University Press, Cambridge MA.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20 :158–177.
- Whitehead, A. N. (1920). *The Concept of Nature*. CUP, Cambridge.
- Williamson, T. (1994). *Vagueness*. Routledge, London.
- Williamson, T. (1997). Précis of vagueness. *Philosophical and Phenomenological Research*, 57 :921–928.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press, Oxford.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Routledge, London. trad fr *Tractatus Logico-Philosophicus* suivi d'*Investigations Philosophiques*, 1961, Gallimard.

- Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Blackwell, Oxford. trad fr *Tractatus Logico-Philosophicus* suivi d'*Investigations Philosophiques*, 1961, Gallimard (pagination de la 3^e édition anglaise/allemande, 2001).
- Wittgenstein, L. (1958). *Blue and Brown Books*. Blackwell, Oxford.
- Wittgenstein, L. (1969). *Philosophische Grammatik*. Blackwell, Oxford. trad fr *Grammaire Philosophique*, 1980, Gallimard.
- Woods, J. (1974). *The Logic of Fiction*. Mouton, La Haye.

Index Nominum

- Ackermann, W., 235
Adams, E. W., 211, 224, 226, 228, 235, 242, 266
Allais, A., 105
Anderson, A. R., 235
Anderson, J. R., 67, 74, 75, 77, 81
Aristote, 2, 21, 172, 305

Baars, B. J., 77, 80
Baddeley, A. D., 74, 77
Barthes, R., 59
Bedford, E., 48, 49
Belnap, N. D., 235
Bennett, J. F., 223
Bergson, H., 115
Block, N., 100

Caillois, R., 27, 28, 30
Carnap, R., 261
Chisholm, R. M., 212
Cowan, N., 77, 83, 112
Currie, G., 17–19, 22, 23, 25, 33, 34, 36, 38, 42–44, 46, 47, 50, 52–57, 60, 66, 67, 69, 74, 177, 178, 193, 205, 213

Davidson, D., 48, 63, 145, 306
Dennett, D. C., 35, 40, 63, 109, 169
Descartes, R., 301
Dretske, F. I., 64
Dupont-Roc, R., 21

Eco, U., 58, 59, 105
Edgington, D., 224, 226, 228, 235
Elster, J., 48

Faulkner, W., 105
Feagin, S. L., 69, 172
Fine, K., 116, 117, 127, 273
Fleming, I., 58
Fodor, J. A., 15, 35, 74, 80, 81
Frege, G., 27, 108

Freud, S., 49
Fuhrmann, A., 224, 226, 227, 235

Gärdenfors, P., 261
Goldie, P., 48, 139
Goldman, A., 67, 69
Goodman, N., 19, 23, 24, 31, 193–195, 211, 223, 225–227, 229, 231, 232, 235, 243, 270, 301, 305
Gordon, R., 67, 69
Graham, G., 39
Grice, P., 306, 344
Gärdenfors, P., 81, 212

Harper, W. L., 212
Harris, P., 67
Hausdorff, F., 122
Heal, J., 67
Hegel, G. W. F., 31
Hilbert, D., 99
Hintikka, J., 165
Holland, J. H., 81
Huizinga, J., 27–29, 31
Husserl, E., 15, 83, 93, 94, 101, 109–112, 121, 130

Ingarden, R., 19, 176–179, 190, 191, 197
Iser, W., 19, 20, 23, 24, 26, 172, 191, 193, 195

James, H., 205
James, W., 48, 83, 93, 111, 112, 121, 145
Johnson-Laird, P. N., 74, 77, 81
Joyce, J., 58
Joyce, J. M., 162, 163

Kafka, F., 53
Kant, E., 134
Kant, I., 301

- Kintsch, W., 67
 Krasner, D. A., 213
 Kraus, S., 311
 Kripke, S. A., 306
 Kuhn, T. S., 97, 134, 135
 Kuratowski, K., 122, 125
 Kvart, I., 235

 Lallot, J., 21
 Lehmann, D., 311
 Leslie, A., 67, 68, 74
 Lewis, D. K., 4–6, 19, 25, 39, 42, 63, 137, 161, 162, 190–193, 203–205, 211–213, 222–227, 231, 232, 235–237, 241–243, 252, 281, 283, 289, 290, 303, 305, 306
 Lowe, E. J., 211, 224, 225, 227

 Magidor, M., 311
 Magnien, M., 21
 Mares, E. D., 224, 226, 227, 235
 Meinong, A., 101

 Nichols, S., 67–70, 74
 Nozick, R., 64

 Parménide, 115
 Parsons, T., 17
 Pavel, T., 59, 184
 Peterson, S. B., 69
 Plagnol, A., 67–69, 74
 Platon, 31
 Popper, K. R., 305
 Potts, G. R., 69
 Pouivet, R., 301
 Putnam, H., 15

 Quine, W. V. O., 63, 68, 108, 301

 Ramsey, F. P., 40–42, 150, 161, 162, 165–168, 212, 305
 Ricoeur, P., 20, 21, 23, 24, 31, 193, 195
 Russell, B., 101
 Ryle, G., 40, 45, 109, 150

 Saint Thomas, 301
 Sambin, G., 125
 Sartre, J. P., 144, 146, 176, 197
 Sartre, J.-P., 143, 144, 146, 178, 179
 Savage, L., 219, 305

 Schacter, D. L., 74, 77
 Schaeffer, J.-M., 17, 29, 54, 66, 70, 170, 173, 176, 178–180
 Searle, J. R., 306
 Sellars, W., 39, 137
 Simons, P., 110
 Solomon, R. C., 48, 49, 145
 Stalnaker, R. C., 81, 211, 212, 225–227, 231, 232, 235, 237, 240, 242, 283
 Stalnaker, R., C., 45
 Stempel, W.-D., 58
 Sterne, L., 192
 Stich, S., 67–70, 74
 Sutherland, W. A., 274

 Todorov, T., 59

 Walton, K. L., 17–19, 22, 23, 25, 26, 30, 33, 34, 36, 38, 42–44, 46, 47, 50, 52–60, 66, 67, 177, 178, 193
 Watson, J. B., 39
 Whitehead, A. N., 125
 Williamson, T., 116, 127, 273
 Wittgenstein, L., 15, 45, 64, 93, 96, 109, 110, 146, 159, 261, 301, 303, 306
 Woods, J., 25

 Zénon, 115

Index Rerum

- « dans la fiction f, ... », 25, 68, 213
- absence, voir *hors sujet*
de l'état instantané de la conscience, 111, voir *focalisé, non focalisé*, gradation *focalisé, non focalisé*, absent, voir localité de l'état instantané de la conscience
- activation, 74
- administrateur central, voir système directorial
- affects relatifs à la fiction, 18, 47
rapports aux États Mentaux usuels, 55
refoulement de, 55, 174, 182
théories de Walton et de Currie, 50
- aire « vague », voir région
- algèbre de Boole, 218, voir conséquence, conception algébrique, voir structure logique, 258
algèbre libre, 258
- algèbre interprétée, voir notion de conséquence, 255, 258
- apparence de réalité, 176
- apport pragmatique, voir pragmatique
appui sur, 237, voir *verum sequitur ad quodlibet*, pertinence
bonne conception de, et rapport à la dynamique, 252
dynamique de, voir dynamique des contextes, 248–249
rapport entre, voir contexte : rapport entre différents, 246–247
- apport pragmatique (à l'analyse de contrefactuels), 225
- apport pragmatique (à l'analyse des contrefactuels), 227
- approche de la fiction
créative, 23
imitative, 22
- approche globale, 300
aux contrefactuels, voir contrefactuels : analyses globales de, voir contrefactuels : approches globales
- approche locale, 300
aux contrefactuels, voir contrefactuels : approches locales
opposition aux approches globales, 302
- approche locale : sol trouvé, 304
- arrière-plan de l'état instantané de la conscience, 118
rapport au triplet *focalisé, non focalisé*, absent, 119
rapport à l'objet et la relation, 110
structure de, 119, voir vague d'état instantané de la conscience, « au bord » et « en profondeur »
- atomisme, 100
- attention, 77
focus attentionnel, 77
- attention scindée, 179
- attribution (des États Mentaux)
base de, 41
bloc de, 39
- behaviorisme, voir fonctionnalisme logique, 39
méthodologique, 39, voir externalisme méthodologique
- bloc d'attribution

- et théorie scientifique, 39
- cadre, 263, voir algèbre interprétée, voir ensemble co-indépendant
- causalité, voir potentialité
- chevauchement, 114, voir espace, 122
- de structures instantanées de la conscience, voir correspondance, 126
- des états instantanés de la conscience, conçus comme régions, 123
- co-indépendance
 - comme ensemble co-indépendant, 260, voir ensemble co-indépendant
 - et indépendance ou non pertinence probabiliste, 261
 - compte tenu du, 262
 - compte tenu du conséquent, 230–232
 - de prémisses et antécédent, 229–230
 - et indépendance mathématique, 260
- compartmentation (d'entités mentales), voir mécanisme spécifique à la fiction, statut
- compétence fictionnelle, 17, 54, 170, voir immersion fictionnelle, voir énoncés portant *sur* un texte de fiction, voir affects relatifs à la fiction, refoulement de, voir lecture, interruption de, voir évitement de confusion entre fiction et réalité, 172–185
- conception ordinaire des États Mentaux, voir État Mental ordinaire
- conception sophistiquée des États Mentaux, voir État Mental sophistiqué
- concrétisation, voir vérité fictionnelle, voir localité et concrétisation
- caractère partielle, 191
- lieux d'indétermination non remarqués, 191
- connexité, 272, voir contexte, définition de
- conscience, voir état instantané de la conscience
 - « irréfléchie » ou « non positionnelle », 143
- approche à partir de, voir stratégie « natuaraliste » en psychologie
- comme aspect psychologique « de base », 33
- comme champ de vision, 93
- comme flux, 93
- dynamique de, voir dynamique de la conscience
- inaccessibilité à l'autrui
 - difficulté d'acquisition, 96
 - difficulté d'approche, 96
- nature de l'enquête
 - théorique plutôt que pratique, 96
- perspective adoptée, 96
- sens de « dans » la conscience, 94
- système spécifique, voir attention
- étendue de, ou sens de, 92
- conscience spécifique à la fiction, 178
- conséquence, notion de, 221, voir contrefactuel, forme générale d'analyse
- analyse en termes d'algèbres interprétées, 254
- comme aspect structurel, 222
- comme structure du contexte, voir contexte : notion de conséquence comme structure de
- conception algébrique, 218
- conception extensionnelle, 218
- deux conceptions (extensionnelle et algébrique)
 - intuitions différentes, 219
 - isomorphisme structurel, 219
- et notion de nécessité, 218
- ligne de partage avec prémisses, voir frontière arbitraire et imprécise

- conséquent
 - conséquent comme équivalent à une combinaison logique des prémisses et de l'antécédent, 262
- contexte, voir pragmatique, voir apport pragmatique
 - application générale au discours, 283
 - apport pragmatique dans l'analyse proposée comme totalité de, 253
 - dynamique de, voir dynamique de contextes
 - définition de, 272
 - notion de conséquence comme structure de, 253
 - rapport entre contextes différents, 278
 - vague de, voir vague du contexte
- continuité, 114, voir dynamique et état instantané, voir espace, topologie
 - comme chevauchement de régions, 122, 284
 - comme contrainte sur la conception d'état instantané, 114, 121
 - et distinction entre aspects diachroniques et synchroniques, 128
 - et effets des aspects *intérieurs* aux structures sur leur dynamique, 129
 - et interprétation des structures instantanées, 136
 - et rapport entre aspects *intérieurs* et *extérieurs* de structures instantanées, 128
 - et vague, 124–127
- contrefactuel, 1, 203, 210
 - analyse d'Adams et d'Edgington
 - pertinence de l'antécédent pour le conséquent, 236
 - analyse d'Adams et d'Edgington (analyse probabiliste), 224
 - analyse de Goodman, 223
 - et l'« indépendance » des prémisses par rapport à l'antécédent, 232
- analyse de Lewis-Stalnaker, 25, 223
 - et l'« indépendance » des prémisses par rapport à l'antécédent, 232
 - pertinence de l'antécédent pour le conséquent, 235
- analyse de Lowe, 224
- analyse de Mares & Fuhrmann, 224
- analyse globale de
 - seesous les théories particulières, 227
- analyse proposée
 - analyse pragmatique *vs* analyse sémantique, 291
 - comme schème d'analyse, 278
 - donnant dynamique comme continue, voir dynamique de contextes
 - retrouve l'analyse de Lewis-Stalnaker par « passage à la limite », 292
 - retrouve sémantique « globale » par « passage à la limite », 291
 - version approfondie, 272
 - version préliminaire, 264
- analyses globales de, 227
- approche locale, 245, 258
- approches globales, voir *sous les théories particulières*
 - faiblesses, 245
- approches globales : manque d'intuition, 245
- comme *connexion*, 211
- difficulté de trouver des propriétés « certainement » logiques, 240, 241
- dépendance de la pragmatique, 225, voir apport pragmatique
- dépendance du contexte, voir contrefactuel : dépendance de la pragmatique
- et *hors sujet*, voir *hors sujet*
- et analyse de vérité fictionnelle,

- voir vérité fictionnelle
- et distinction conditionnel subjonctif / conditionnel indicatif, 211
- et révision de croyances, voir croyance
- forme générale d'analyse, 221, voir conséquence, prémisses, voir contrefactuel, analyse de Goodman, de Lewis-Stalnaker, de Mares & Fuhrmann, de Lowe, d'Adams et d'Edgington
- compréhension dans la conception algébrique, 222
- compréhension dans la conception extensionnelle, 221
- importance de règles logiques, 244
- parallélisme avec la fiction, 203, 212, voir contrefactuel : rapport à la réalité
- part non formelle, voir apport pragmatique
- pragmatique de, voir apport pragmatique, pragmatique
- importance de, 243
- problème « logique » de, voir contrefactuels : problème « sémantique »
- problème « pragmatique », voir pragmatique, 241, voir apport pragmatique
- comme le problème de contrefactuels, 244, voir contrefactuels : la question de, rapport à la réalité, théorie complète de
- problème « sémantique », voir sémantique, logique
- facilité, manque d'intérêt de, 244
- problème
 - « sémantiquepragmatique », 241
- question de, 214, voir contrefactuel : rapport à la réalité
- comme descriptive, 215
- comme générale, 214
- comme méthodologique, 215
- comme question de dynamique d'apport pragmatique, voir apport pragmatique : dynamique de
- liste de défis, 249–250
- méthode de ramasser des « données », 215
- rapport à la réalité, 203, 212
- théorie complète de
 - comme théorie du bloc sémantique-pragmatique, 244
- correspondance, voir interprétation des états instantanés
- difficulté de, entre structures instantanées de la conscience à des instants éloignés dans le temps, 105
- entre structures instantanées de la conscience, 104, voir localité, totalité de la structure de la conscience
- rapport à la dynamique, 126, voir dynamique, continuité
- entre un objet et une relation d'une structure, 101, 107
- croyance, 5, voir *faire comme si*, 36, voir État Mental usuel, voir expérience phénoménale et croyances, voir potentialité
- « activité », 55
- « effective » ou « en jeu », 150
- « fictionnelle », 17
- activité et présence dans la structure instantanée, 155, voir croyance : corrélation avec aspects des structures instantanées
- comme État Mental narratif, 45
- comme propriété de succession de structures instantanées de la conscience
- « raffinement » proposé du terme « disposition », 163
- comme propriété des succes-

- sions des structures instantanées de la conscience, 151
 - comme tendance, 40
 - degré de, 41
 - revision de, 212
 - tentative ramseyenne de « mesurer », 165
 - tentative ramseyenne de « mesurer », 168
 - terminologie employée, 138, voir Etat Mental : neutralité, 150
- croyance fictionnelle, 13
- croyances
 - corrélation avec aspects des structures instantanées à des instants particuliers, 152
- dire, 108, 109
- distension, 268, voir cadre
 - « logique » *vs* « non logique », 270
 - de deux cadres dans une algèbre de Boole commune, 280
- dynamique, 6
 - contraintes méthodologiques relatives à, voir continuité, 121–123
 - d'apport pragmatique, 248–249
 - comme question de contre-factuels, 248
 - de contextes, 282–286
 - comme continue, 282–286
 - de l'expérience phénoménale, 57
 - de la conscience, 94
 - possibilité de l'aborder dans une structure instantanée particulière, 129, voir continuité et rapports entre aspects *intérieurs* et *extérieurs* de la structure instantanée
 - des états instantanés de la conscience, voir espace, topologie, continuité, chevauchement, états instantanés comme régions
 - du rapport entre États Mentaux spécifiques à la fiction et États Mentaux usuels, 55
 - défi de, 113
 - et « forme » de la structure instantanée, 130
 - et *totalité* de la structure instantanée, voir dynamique et « forme » de la structure instantanée
 - et effets des aspects des structures instantanées, 129
 - et l'état instantané, 91
 - et mécanisme central, 75
 - et théorie de la fiction, voir fiction, théorie de
 - et état instantané, voir défi de la dynamique, voir continuité, 128
 - lecture comme, 172
- émotion, 36, 46–51
 - théorie « cognitive » de, voir thèse de sentiment comme configuration de pensées
 - théorie « jamesienne » de, voir thèse de sentiment comme aspect *sui generis*
 - théorie « phénoménologique » de, voir thèse de sentiment comme aspect global
- énoncés portant *sur* un texte de fiction, 173, 181, 193
- énoncés transfictionnels, 193
- enracinement épistémique, 305
- ensemble co-indépendant, 260, voir co-indépendance, voir prémisses
 - incompatibilité de plusieurs, 261
- entité mentale, 67
- épistémologie, 305, voir potentialité
- epochè*, voir perspective « internaliste »
- epochè*, 15
- espace, 114, voir topologie, continuité, 122, 124–125
 - et région, et chevauchement, 122

- pour dynamique des états instantanés de la conscience, 123
- état instantané de la conscience, 6, 94, voir structure instantanée de la conscience
- arrière-plan, premier plan, voir arrière-plan, premier plan de l'état instantané de la conscience
- chevauchement de, voir chevauchement des états instantanés
- comme régions, 123
- conception de, voir continuité comme contrainte
- dynamique de, voir dynamique de la conscience
- et théorie de la fiction, voir fiction, théorie de
- interprétation de, voir correspondance, voir établir état instantané comme « données », 133–136
- et continuité, 136
- et totalité de la structure, 134
- localité de l'interprétation, 135, voir État Mental : base d'attribution locale
- non unicité de l'interprétation, 135
- rapport avec correspondance, 134
- sa localité, voir localité de l'état instantané de la conscience
- son vague, voir vague d'état instantané de la conscience
- établir comme « données » pour une théorie, 95–98
- étendue de la notion, 83
- État Mental
 - attribution de, voir attribution (des États Mentaux)
 - base d'attribution
 - comportant structures instantanées de la conscience, 137
 - base d'attribution globale, 42
 - et le *hors sujet*, 162
 - base d'attribution locale, 41
 - dans le cadre proposé, 137, voir pensée, sentiment, croyance
 - distinction narratif / non narratif dans cadre « internaliste », 138
 - narratif, 44–46
 - dans le cadre proposé, voir croyance, 148–168
 - neutralité à l'égard de thèses réalistes, instrumentalistes, *etc.*, 138
 - non narratif, 44–46
 - non saturé, 41
 - ordinaire, 38–44
 - phénoménologiquement
 - présent, 45, 47
 - rapport complexe entre, 198
 - saturé, 41
 - et le *hors sujet*, 162
 - sophistiqué, 38–44, 63, 64
 - « raffinement » proposé du terme « disposition », 164
 - et l'étourderie, 164
 - privilège la précision à la compréhension de la dynamique, 163
 - écrase le *hors sujet*, 162
 - spécifique à la fiction, 38, 43, voir *faire comme si*, 52–60
 - usuel, 38, 60–65
 - variabilité de modalités, 64
 - volonté non révisionniste à l'égard de, 137
 - vs* mécanisme, 35–37
- évitement de confusion entre fiction et réalité, 174, 183–185
- explicite, 109, voir tacite
- expérience phénoménale, voir dynamique de l'expérience phénoménale, voir état instantané de la conscience
- du lecteur, 56
- et croyances, 64
- et mécanisme central, 76, voir attention
- externalisme
 - relatif au contenu des États Mentaux, 86
- externalisme méthodologique, 39

- et la distinction Etat Mental
narratif *vs* non narratif, 45
- faire comme si*, 5, 22, 23, 36, voir
Etat Mental spécifique à la
fiction, 42–44
- comme Etat Mental narratif,
46
- conception de, 42–44
- désirs dans le jeu de, 42
- et croyance, 22
- et théorie de la fiction, voir fic-
tion, théorie de
- intuitions et le *hors sujet*, 177,
voir localité
- jeu de, 22
- rapport aux croyances,
53–55, voir expérience
phénoménale du lecteur
- variabilité des modalités, 58
- faire comme si*
et vérité fictionnelle, 22
- feinte, voir *faire comme si*
- mécanismes relatives à la, 66
- fiction
 - apparition *focalisée* dans la
structure instantanée de la
conscience, 181–183
 - apparition *non focalisée* dans la
structure instantanée de la
conscience, 183–185
 - entités fictionnelles, 17, 193
 - et moral, 31
 - parallélisme avec le contrefac-
tuel, 203, voir contrefac-
tuel
 - question de, 14
 - question psychologique, voir
psychologie du lecteur
 - question sémantique de la fic-
tion, 25
 - rapport à la réalité, voir rap-
port fiction et réalité
 - refigurer la réalité, voir ap-
proche créative de la fic-
tion, voir rapport fiction et
réalité, effets de la fiction
 - termes fictionnels
 - application à la réalité, 19,
voir rapport fiction et
réalité, effets de la fiction,
 - 194
- théorie de
 - États Mentaux spécifiques
à la fiction, voir Etat
Mental, rapport complexe
entre
 - état instantané de la
conscience, 170, 172
 - États Mentaux spécifiques à
la fiction, 171, 196–199
 - la dynamique, 171, 186
 - stratégie pour développer,
169–172
 - théorie de la psychologie de, 47
- fidélité, maxime de, 97, voir neutra-
lité d'outils et de notions
- focalisé*, 106, voir objet, voir nom,
dire, explicite, manifeste,
voir *non focalisé*, absence,
gradation *focalisé*, *non fo-
calisé*, absent
- fonction, voir prédicat
- fonctionnalisme, 39, voir behavio-
risme
- forme d'argumentation, voir logique
- frontière arbitraire et imprécise
entre notion de
conséquence et prémisse,
227
- dans l'analyse proposée, 268,
voir restriction, distension
- fécondité, voir fécondité de
mécanisme
- des États Mentaux sophis-
tiqués, 63
- du mécanisme central pour
penser la fiction, 76
- fécondité, maxime de, 97
- gradation *focalisé*, *non focalisé*, ab-
sent, 112
- holisme, 100
- hors sujet*
 - et absence
 - de l'algèbre, 239
 - de l'état instantané de la
conscience, 111, voir ab-
sence, voir absence
 - et contrefactuels, 238
 - et croyance, 62, voir croyance :
« activité »

- et indépendance, 63, 239
- et l'immersion fictionnelle, 54, 175
- et mécanismes, 80
- et structures instantanées de la conscience, voir localité de la structure instantanée de la conscience
- écrasé par État Mental sophistiqué, voir État Mental sophistiqué, saturé, base d'attribution globale
- hypothèse d'états de la conscience exposés, 97
 - est d'ordre méthodologique, 97
- hypothèse de connaissance du fait de fiction, 16, 169
- identification
 - entre structures instantanées de la conscience, voir correspondance entre structures instantanées de la conscience
- imagination, voir *faire comme si*
- imagination voir *faire comme si* 23
- immersion fictionnelle, 17, 173,
 - voir apparence de réalité,
 - voir conscience spécifique à la fiction, voir attention scindée, 175–180
- et le *hors sujet*, 175
- implication
 - matérielle, 218, voir conséquence, conception de
 - comme aspect matériel, 220, voir dire, manifeste
 - stricte, 218, voir conséquence, conception de
 - comme aspect structural, 220, voir montrer, manifeste
- interprétation des structures instantanées de la conscience
 - double importance, 133
- investissement dans la lecture, voir apparence de réalité
- jeu, 26–32
 - contraintes réelles, voir fond réel
- espace abstrait, 28–30
- espace de, voir jeu, région de, 70
- fond réel, 27–32
- liberté, 27
- nivellation, 30
- région de, 29
- réglementation, 27, 28
- séparation, 28
- langage local, 255
- lecture
 - comme démarche dynamique, voir dynamique, lecture comme
 - interruption de, 54, 174, 182
- lieu d'indétermination, 190, voir concrétisation
- localité, voir globalité, voir site, voir *hors sujet*
 - de l'état instantané de la conscience, 103
 - de la structure instantanée de la conscience, 103–105
 - et concrétisation, 191
 - rapport intérieur et extérieur de *faire comme si*, 178
- logique, voir sémantique
 - confusion avec généralité, 243
 - normative *vs* descriptive, 215, 245
- make-believe, voir *faire comme si*
- make-desire, voir désir dans le jeu de *faire comme si*
- manifeste, 109
- mimésis*, 21
- mode de sélection, 58
- mondes possibles, 4, voir relation de similitude, voir conséquence, conception extensionnelle de
 - retrouvés à partir des cadres, 291
- montrer, 109
- mécanisme, voir simulateur, voir entité mentale
 - central, voir activation, voir système directorial, voir attention

- et fiction, voir fécondité du mécanisme central pour penser la fiction
- gestion de la dynamique, voir dynamique du mécanisme central
- fécondité, 37, 65
- ordinaire, 79–82
- réalisation physique / opérabilité informaticque, voir fécondité de mécanisme
- spécifique à la fiction, voir expérience phénoménale et mécanisme central
- distinction par processus, 69
- distinction par statut, 68
- limites, 71–78
- présentation, 66–71
- rapport à mécanismes ordinaires, 72–75
- statut *vs* processus, 67
- théorie hybride, 70
- vs* État Mental, 35–37
- mécanisme : spécifique à la fiction
 - recours au mécanisme central, voir mécanisme central
- mécanismes
 - spécifiques à la fiction
 - variabilité de modalités, 77, voir *faire comme si*, variabilité de modalités
- métaphysique, 304, voir méthodologie, voir potentialité
- méthodologie, voir hypothèse d'états de la conscience exposés
- et métaphysique, ontologie, épistémologie, 301
- économie méthodologique, 15
- neutralité, voir structure et méréologie, et logique d'outils et de notions, 98, 100
- nom, voir dire, parler de, parler explicitement, 107–109
- non focalisé*, 106, voir relation, voir prédicat, montrer, tacite, voir *focalisé*, absence, gradation *focalisé*, *non focalisé*, absent
- noyau sémantique, voir sémantique, pragmatique : limite entre pragmatique et sémantique
- nécessité, notion de, voir conséquence, notion de
- objet, voir structure, relation, voir neutralité d'outils et de notions, 98–102, voir *focalisé*, voir nom, voir explicite, voir manifeste
 - relation avec premier plan, arrière-plan, voir premier plan, arrière-plan
- ordre d'analyse
 - vs* ordre d'exposition, 5, 92
- ordre d'exposition, voir ordre d'analyse
- ouvert, voir topologie, voir vague
 - collection ouverte, 116
 - collection ouverte de base, 274
 - et dynamique, 124–127
 - structure ouverte, 116
- parler de, 109
- parler explicitement, 108
- pensée, 139–143
 - « être conscient » de ses propres, 142
 - focalisée, 141
 - non focalisée, 141
- perspective « globaliste »
 - et différence entre cadres_f et modèles préférentiels, 327
- perspective « internaliste », 15
 - conséquences pour approche de la fiction, 83
 - différence de l'*epochè*, 94
 - opposition à l'externalisme méthodologique, 83
- perspective internaliste, 39
- pertinence de l'antécédent pour le conséquent, voir *verum sequitur ad quodlibet*, voir semifactuels indépendants *vs* non indépendants, 232–238
- philosophie du langage, 305, voir potentialité

- potentialité, 301–306
 notion générale de, 301
- pragmatique
 comme problème de contrefactuels, 244, voir contrefactuels : rapport à la réalité, problème « pragmatique »
 couche pragmatique, 237
 importance de apport pragmatique
 appui sur, 243
 pragmatique engagée, 236
 relation entre pragmatique et sémantique
 complémentarité de, 242
 relation entre pragmatique et sémantique : limite difficile à tracer, 241
- premier plan de l'état instantané de la conscience, 118
 rapport à l'objet et la relation, 110
- propension, voir potentialité
- proximité, voir espace
- prédictat, voir montrer, 107–109
- prémisse, 221, voir contrefactuel, forme générale d'analyse, voir contrefactuel, analyse de Lowe, d'Adams et Edgington
- « indépendance » par rapport à l'antécédent, voir co-indépendance de prémisses et antécédent
- « indépendance » par rapport à l'antécédent compte tenu du conséquent, voir co-indépendance compte tenu du conséquent
- analyse en termes des ensembles co-indépendants, 260–263
- comme aspect matériel, 222
- ligne de partage avec notion de conséquence, voir frontière arbitraire et imprécise
- psychologie
 question causale, voir mécanisme
 question conceptuelle, voir État Mental
- psychologie du lecteur, 5, 14
- quasi émotion, 50
- quasi-jugements, voir conscience spécifique à la fiction
- question de l'induction, voir potentialité
- rapport fiction et réalité, 1, 13, 16–20
 apports de la réalité, voir vérité fictionnelle, 187–192
 chevauchement, 18, 30, 185–196
 effets de la fiction, 19, 31, 192–196
 question complexe et difficile, 18
 similitude et distinction, 16, 22
 transferts, voir chevauchement
- rapport symbiotique entre l'état instantané et la dynamique, voir dynamique et état instantané
- relation, voir structure, objet, voir neutralité d'outils et de notions, 98–102, voir *non focalisé*, voir prédicat, ou fonction, voir manifeste, voir tacite
- relation avec premier plan, arrière-plan, voir premier plan, arrière-plan
- relation de similitude, voir mondes possibles
- relation de similitude comparée entre mondes, voir contrefactuels, analyse de Lewis-Stalnaker
- restriction, 268, voir cadre
 « logique » *vs* « non logique », 270
- règle logique, voir logique
- réalité
 conception de, 23
- région, 124–125
- savoir (par opposition à *savoir faire*), 109, voir explicite
- savoir faire*, 109, voir tacite
- semifactuel, 211

- distinction semifactuels
 - indépendants *vs* non indépendants, 234
 - dans l'analyse proposée, 277
- sentiment
 - comme aspect global, 146, voir structure instantané de la conscience : « forme » de et la dynamique, 147, voir dynamique et « forme » de la structure instantanée
 - comme Etat Mental non narratif, 45
 - et correspondance à une configuration d'objets et relations, 144
 - non narratif, 46–51
 - thèse de sentiment comme aspect *sui generis*, 145
 - thèse de sentiment comme configuration de pensées, 145
 - éprouver sans tenir compte qu'on l'éprouve, 144
- sentiment de réalité, voir apparence de réalité
- simulateur, 36
- simulation, voir mécanisme spécifique à la fiction, processus
- site, 6, 300, voir localité, structure instantanée de la conscience, cadre, contexte
- sol, 300
 - de possibilités, 303
- solipsisme méthodologique, 15
- stratégie « mimétique », 21–26, 32
 - dans la perspective « internaliste », 84
 - en matière de « vérité fictionnelle », contrefactuels, 204
- stratégie « naturaliste », 32
 - en matière des contrefactuels, 207
 - en psychologie du lecteur, 85–87
- structure, voir objet, relation, voir neutralité d'outils et de notions, 98–102
 - et logique, 110
 - et méréologie, 110
 - ouverte, voir ouvert
- structure instantanée de la conscience, 101, voir état instantané de la conscience
 - « forme » de, voir dynamique et « forme » de la structure instantanée, voir sentiment comme aspect global
 - comme base d'attribution des États Mentaux, voir Etat Mental : base d'attribution
 - ouverte, voir arrière-plan, premier plan de l'état instantané de la conscience, voir ouvert
 - sa localité, voir localité de la structure instantanée de la conscience
 - sa totalité, voir totalité de la structure instantanée de la conscience
 - totalité de, voir structure instantanée, « forme » de
- structure supplémentaire, voir sol
- structures instantanées de la conscience
 - chevauchement, voir chevauchement, correspondance
- système de pensée, voir réalité, conception de, 195
- système directorial, 74
- sémantique, voir logique
 - vs* pragmatique, voir pragmatique
 - confusion avec généralité, 243
 - noyau sémantique
 - instabilité de, 241
- tacite, 109, voir explicite
- théorie « simulationniste », voir mécanisme spécifique à la fiction, processus, voir simulateur
- théorie « théorique », voir mécanisme spécifique à la fiction, statut

- théorie de la décision, voir État Mental sophistiqué
- théorie globale, 226
- topologie, 122, voir espace, chevauchement, continuité, 124–125, voir ouvert
- deux approches, voir espace topologique, et topologie formelle
- espace topologique, 125
- topologie formelle, 125
- totalité
- de la structure instantanée de la conscience, 101, 104
- vague, voir ouvert
- approche visant à comprendre la *structure* du, 273, voir vague et topologie
- d'état instantané de la conscience, 116
- « au bord », 119
- « en profondeur », 119
- du contexte, 272, voir frontière imprécise entre notion de conséquence et prémisses
- est structuré, 273
- et continuité, 124–127
- et dynamique, 124–127
- et topologie, 273, voir topologie, ouvert
- fécondité d'approche du, 117, voir vague et dynamique
- méthodologie de l'approche du, 117
- théorie supervaluationniste de, 117
- théorie épistémique de, 116
- verum sequitur ad quodlibet*, 234, voir pertinence de l'antécédent pour le conséquent, voir semifactuels indépendants
- réponse au problème de
- ayant recours à la couche pragmatique, 237
- ayant recours à la pragmatique engagée, 237
- en le dévalorisant, 236
- verum sequitur ad quodlibet* dans l'analyse proposée, 277
- vérité fictionnelle, 4, 19, voir *faire comme si*, 186, 190, voir « *dans la fiction f, ...* »
- analyse en termes de contrefactuels, 25, 204, 213
- n'est pas analyse en termes de mondes possibles, 204
- analyse en termes des contrefactuels, 214
- application aux textes non fictionnels, 233
- analyse lewisienne en termes de mondes possibles, 204, voir stratégie « mimétique »
- dépendance du contexte, 205, voir apport pragmatique
- et contrefactuels, 204
- problème d'interprétations multiples, 205
- vérités *autorisées* par le texte de fiction, 234
- Zénon, paradoxe de, 115

Table des matières

Avertissement au lecteur	iii
Introduction	1
 I Jouer à croire : l'examen des processus mentaux impliqués dans la lecture des textes de fiction	 9
1 Attitudes à l'égard de la fiction	11
1.1 La problématique de la fiction	11
1.1.1 Le seuil de la problématique : quatre questions	11
1.1.2 Le contour de la problématique	13
1.1.3 Le cœur de la problématique	16
1.2 Approches générales de la fiction	20
1.2.1 La stratégie « mimétique »	21
1.2.2 Le jeu et la fiction	26
1.2.3 La stratégie adoptée	32
1.3 Approches psychologiques	34
1.3.1 La question « conceptuelle » et la question « causale » . .	35
1.3.2 Conceptions ordinaires et sophistiquées des États Mentaux	37
1.3.3 États mentaux narratifs et non narratifs	44
1.3.4 Émotions et sentiments	46
1.4 États Mentaux spécifiques à la fiction	51
1.4.1 Difficultés pour la théorie des États Mentaux spécifiques à la fiction	52
1.4.2 Difficultés des États Mentaux usuels	60
1.5 Mécanismes spécifiques à la fiction	65
1.5.1 Quelques mécanismes psychologiques	66

1.5.2	Les limites des théories proposant des mécanismes spécifiques à la fiction	71
1.5.3	Limites des mécanismes usuels	78
1.5.4	Repartir de la conscience	82
1.6	Résumé : question et stratégies	87
2	La psychologie du complexe	89
2.1	Faire face à la conscience	91
2.1.1	La conscience : intuitions et enjeux	92
2.1.2	Analyser l'état instantané de la conscience : introduction	95
2.1.3	Notions pour l'analyse de l'état instantané de la conscience : les notions de structure, d'objet et de relation	98
2.2	L'état instantané de la conscience	102
2.2.1	La localité de la structure instantanée de la conscience . .	102
2.2.2	Objets et relations : le focalisé et le non focalisé	105
2.2.3	Résumé : le triplet focalisé / non focalisé / absent	111
2.3	La dynamique de la conscience	113
2.3.1	Le caractère vague de la structure	115
2.3.2	La continuité de la dynamique	121
2.3.3	La continuité et le vague	123
2.3.4	Rapport entre la structure instantanée de la conscience et sa dynamique	128
2.3.5	Résumé : dynamique, continuité et vague	131
2.4	Le rôle de l'interprétation	132
2.5	Retrouver les États Mentaux	136
2.5.1	Les pensées	140
2.5.2	Les sentiments	143
2.5.3	États Mentaux narratifs	148
3	Retrouver la fiction	169
3.1	Compétence fictionnelle	172
3.1.1	L'état d'immersion fictionnelle	175
3.1.2	Les mises en jeu du fait de fiction	180
3.1.3	Confusions évitées	183
3.2	Chevauchement entre fiction et réalité	185
3.2.1	Question C. : L'irruption de la réalité	186
3.2.2	Question D. : L'infiltration de la fiction	192
3.3	États Mentaux spécifiques à la fiction	196

II Jeux entrecroisés :	
fiction et le problème des contrefactuels	201
Introduction	203
4 Déterrer le problème	209
4.1 Rapport à la fiction	210
4.1.1 Contrefactuels	210
4.1.2 Fiction et contrefactuels	211
4.1.3 La question des contrefactuels	214
4.2 Intuitions, données et défis	216
4.2.1 Intuitions traditionnelles	217
4.2.2 La frontière imprécise et arbitraire	227
4.2.3 Indépendance et pertinence	228
4.2.4 Le <i>hors sujet</i>	238
4.3 La question « difficile » des contrefactuels	240
4.3.1 La question des contrefactuels : sémantique et pragmatique	240
4.3.2 Défis externes : comparaison d'apports pragmatiques entre situations différentes	246
4.3.3 Défis externes : la dynamique	248
4.3.4 Résumé : la question des contrefactuels	249
5 Aborder les contrefactuels	251
5.1 Une théorie des contrefactuels	252
5.1.1 La notion de conséquence et l'algèbre interprétée	253
5.1.2 La prémisse et la co-indépendance	259
5.1.3 Une analyse préliminaire des contrefactuels	263
5.1.4 La frontière imprécise et arbitraire et le rapport entre cadres	267
5.1.5 Définition du contexte	271
5.2 La théorie face aux défis	275
5.2.1 Conséquences immédiates	276
5.2.2 La comparaison de contextes	278
5.2.3 La dynamique de contextes	282
5.2.4 Résumé : Bilan de la théorie proposée	286
5.3 Conséquences de la théorie	288
5.3.1 Flexibilité de la partition entre sémantique et pragmatique	289
5.3.2 Concevoir la logique	292
Conclusions et Projections	299

A La logique locale	309
A.1 La logique non monotone aujourd'hui	311
A.2 Règles comme propriétés des cadres	313
A.2.1 Préliminaires : une analyse relâchée	316
A.2.2 L'analyse (ALP) en termes d'ordres	317
A.2.3 Règles pour les analyses proposées	320
A.3 Règles et transitions entre cadres	324
A.3.1 Préliminaire : le cadre _f	325
A.3.2 Règles comme opérations : Séquents et cadres _f	328
A.3.3 Règles comme opérations : Fusions, produits et quotients des algèbres	329
A.3.4 Règles comme opérations : Fusions, produits et quotients des cadres _f	335
A.3.5 Les règles sur les cadres	345
A.4 Règles comme contraintes globales	347
A.4.1 Qu'est qu'une sémantique globale	348
A.4.2 « Sémantique globale » comme cadre gonflé	349
Bibliographie	359
Index Nominum	373
Index Rerum	375

Jouer avec le faux

Résumé : Il est bien connu que la fiction, alors qu'elle est *distincte* de la réalité, entretient avec elle des *similitudes* importantes. Quant aux contrefactuels, leur rapport à la réalité est lui aussi double : ils renvoient, certes, à des situations *irréelles*, mais aux situations irréelles les plus *similaires* à la réalité. Cette perspective dominante repose cependant sur la notion de similitude, laquelle est non seulement évasive, mais variable. Une analyse plus poussée du rapport entre la fiction et la réalité d'une part, et entre les contrefactuels et la réalité d'autre part, est-elle possible ? Ce travail offre un cadre conceptuel qui permet une approche approfondie de ces rapports, en mettant l'accent sur leur *dynamique*. En développant, d'abord dans le cas de la psychologie de la fiction, ensuite dans le cas des contrefactuels, une représentation fine de l'état de ce rapport à un instant particulier, et une théorie de son changement, on se donne des outils pour comprendre non seulement les détails du rapport entre la fiction (respectivement les contrefactuels) et la réalité à un moment donné, mais également le développement antérieur qui a mené à la situation où ce rapport, avec ces propriétés, est à l'œuvre.

Gambling with falsity

Summary : It has oft been noted that, whilst fiction is *distinct* from reality, there are important *similarities* between the two. This double-edged relationship to reality is also present in the case of counterfactuals : although they refer to *non actual* situations, the pertinent non actual situations are those which are most *similar* to reality. However, this dominant perspective relies heavily on the notion of similarity, which is not only elusive, but varies according to the circumstance. Is there a more detailed analysis to be had of the relationship between fiction and reality on the one hand, and counterfactuals and reality on the other ? This dissertation proposes a conceptual framework which offers an enhanced understanding of these relationships, notably by way of an understanding of their *dynamics*. The development, firstly in the case of the psychology of fiction, and then in the case of counterfactuals, of a careful representation of the state of the relationship at a particular instant, and of a theory of its dynamics, will furnish the tools required not only for a detailed analysis of the relationship between fiction (respectively counterfactuals) and reality at a particular moment, but furthermore for an account of the progression by which this particular relationship, with these particular properties, came into effect.

Discipline : Philosophie

Mots clefs : Fiction ; immersion fictionnelle ; vérité fictionnelle ; *faire comme si* ; philosophie de l'esprit ; théorie de la conscience ; contrefactuels ; contexte ; sémantique ; pragmatique ; logique philosophique ; dynamique ; localité.

Équipe d'accueil : IHPST

13, rue du Four
75006 Paris.